Université Mohammed V - Agdal
Faculté des Sciences
Département de Mathématiques et Informatique
Avenue Ibn Batouta, B.P. 1014
Rabat, Maroc

Filière :
Sciences Mathématiques et Informatique (SMI)
et
Sciences Mathématiques (SM)

Module Analyse Numérique I :

Par

Année 2003-2004

Plan du cours

Introduction

Représentation des nombres en machine

Résolution de f(x)=0
- **Méthodes directes**
- **Méthodes itératives**

Résolution de systèmes linéaires
- **Par la méthode de Gauss**
- **Par la décomposition LU**

Interpolation polynomiale

Intégration - Dérivation

Equations différentielles

Chapitre I :

Représentation des nombres en machine

0.1

Arithmétique des calculateurs et Sources d'erreurs

Si sophistiqué qu'il soit , un calculateur ne peut fournir que des réponses approximatives. Les approximations utilisées dépendent à la fois des contraintes physiques (espace mémoire, vitesse de l'horloge...) et du choix des méthodes retenues par le concepteur du programme . (pour plus de détails sur le fonctionnement d'un ordinateur et la terminologie de base voir par exemple la page web htttp://www.commentcamarche.com

Le but de ce chapitre est de prendre connaissance de l'impact de ces contraintes et de ces choix méthologiques. Dans certains cas il doit être pris en compte dans l'analyse des résultats dont une utilisation erronée pourrait être coûteuse.

La première contrainte est que le système numérique de l'ordinateur est discret, c'est à dire qu'il ne comporte qu'un nombre fini de nombres; Il en découle que tous les calculs sont entachés d'erreurs.

0.1.1 Evaluation de l'erreur

Rappelons d'abord quelques notion de base ;

Si X est une quantité à calculer et X^* la valeur calculée, on dit que :

1. $X - X^*$ est l'erreur et $\mid E \mid = \mid X - X^* \mid$est l'erreur absolue.

   Exemple :

   Si $X = 2.224$ *et* $X^* = 2.223$ *alors l'erreur absolue* $\mid E \mid = \mid X - X^* \mid = 2.224 - 2.223 = 0.001$

2. $E_r = \left| \frac{X - X^*}{X_r} \right|$ est l'erreur relative, $X_r \neq 0$. X_r est une valeur de référence pour X. En général ,on prend $X_r = X$.

   Exemple :

   Si $X = 2.224$ *et* $X^* = 2.223$ *alors , si on prend* $X_r = X$ *, l'erreur relative* $E_r = \left| \frac{X - X^*}{X_r} \right| = \frac{|X - X^*|}{|X|} = \frac{0.001}{2.224} = 4.496 \times 10^{-4}$

   Cependant, si X est la valeur d'une fonction $F(t)$ avec $a \leq t \leq b$, on choisira parfois une valeur de référence globale pour toutes les valeurs de t.

   Exemple :

   Si $X = \sin(t)$ *avec* $0 \leq t \leq \frac{\pi}{4}$, *on pourra prendre* $X = \frac{\sqrt{2}}{2} = \sup_{0 \leq t \leq \frac{\pi}{4}} \sin(t)$.

En général , on ne connait pas le signe de l'erreur de sorte que l'on considère les erreurs absolues et les erreurs relatives absolues.

Les opérations élémentaires propagent des erreurs.

Dans la pratique, on considère que :

1) L'erreur absolue sur une somme est la somme des erreurs absolues.

2) L'erreur relative sur un produit ou un quotient est la somme des ereurs relatives.

On peut estimer l'effet d'une erreur E sur l'argument x d'une fonction $f(x)$ au moyen de la dérivée de $f(x)$. En effet $f(x+E) \simeq f(x) + Ef'(x)$

Exemple :

Calculer la valeur de $(11111111)^2$

La valeur fournie par une petite calculatrice à cinq chiffres est $1,2345x10^{14}$

Mais la réponse exacte est 123456787654321*.*

La machine a donc tronqué le résultat à 5 chiffres et l'erreur absolue est de $6 * 19^9$*.*

L'erreur relative est de 0.0005% *.*

Cet exemple montre qu'il faut établir clairement l'objectif visé.

Cet objectif est double ;

1) Nous voulons un bon ordre de grandeur (ici 10^{14}) et avoir le maximum de décimales exactes,

2) Ce maximum ne peut excéder la longueur des mots permis par la machine et dépend donc de la machine

0.1.2 La mémoire de l'ordinateur : le stockage des nombres

La mémoire d'un ordinateur est formée d'un certain nombre d'unités adessables appelées OCTETS . Un ordinateur moderne contient des millions voir des milliards d'octets. Les nombres sont stockés dans un ordinateur comme ENTIERS ou REELS.

Les nombres entiers :

Les nombres entiers sont ceux que l'on utilise d'habitude sauf que le plus grand nombre représentable dépend du nombre d'octets utilisés:

-avec deux (2) octets, on peut représenter les entiers compris entre

$$-32768 \text{ et } 32767$$

-avec quatre (4) octets on peut représenterr les entiers compris entre

$$-2147483648 \text{ et } 2147483647$$

Les nombres réels

Dans la mémoire d'un ordinateur, les nombres réels sont représentés en notation flottante.

Cette notation a été introduite pour garder une erreur relative à peu prés constante; quelque soit l'ordre de gandeur du nombre qu'on manipule.

En notation flottante, un nombre a la forme:

$$x = \pm Y \times b^e$$

b est la base du système numérique utilisé

Y est la mantisse : une suite de s entier $y_1 y_2 ... y_s$ avec $y_1 \neq 0$ si $x \neq 0$ et $0 \leq y_i \leq (b-1)$

e est l'exposant(un nombre entier relatif)

La norme choisie est celle où la mantisse est comprise entre 0 et 1 et où le premier chiffre après la virgule est différent de zéro.

Calcul de l'erreur

Nous terminons ce chapitre en définissant les notions de troncature et d'arrondie.

Exemple :

En base 10, $x = 1/15 = 0.066666666......$

Dans le cas d'une représentation tronquée nous aurons, pour $s = 5$, $fl(x) = 0.66666 * 10^{-1}$.

Remarquez comment nous avons modifié l'exposant afin de respecter la règle qui veut que le premier chiffre de la mantisse ne soit pas nul .

Dans ce cas, l'erreur absolue $X - fl(X)$ est de 6×10^{-7}. L'erreur relative est de l'ordre de 10^{-5}

Dans une représentation tronquée à s chiffres, l'erreur relative maximale est de l'ordre de 10^{-s}

Dans une représentation arrondie, lorsque la première décimale négligée est supérieure à 5, on ajoute 1 à la dernière décimale conservée.

Exemple :

$x = 1/15 = 0.066666666$.

Nous écrirons $fl(x) = 0.66667 \times 10^{-1}$

L'erreur absolue serait alors 3.333×10^{-7} *et l'erreur relative serait* 5×10^{-6}

En général, l'erreur relative dans une représentation arrondie à s chiffres est de $5 \times 10^{-(s+1)}$ soit la moitié de celle d'une représentation tronquée.

0.2 Les régles de base du modèle

Pour effectuer une opération sur deux nombres réels, on effectue l'opération sur leurs représentations flottantes et on prend ensuite la représentation flottante du résultat.

l'addition flottante

$$x \oplus y = fl(fl(x) + fl(y))$$

la soustraction flottante

$$x \ominus y = fl((x) - fl(y))$$

la multiplication flottante

$$x \otimes y = fl(fl(x) \times fl(y))$$

la division flottante

$$x \div y = fl(fl(x)/fl(y))$$

Chaque opération intermédiaire dans un calcul introduit une nouvelle erreur d'arrrondi ou de troncature.

Dans la pratique, il faudra se souvenir du fait que deux expressions algébriquement équivalentes peuvent fournir des résultats différents et que l'ordre des opérations peut changer les résultats.

Pour l'addition et la soustraction on ne peut effectuer ces 2 opérations que si les exposants sont les mêmes. On transforme le plus petit exposant et

donc on ne respecte plus la régle voulant que le premier chiffre de la mantisse ne soit pas nul.

Quelques remarques sur ce modèle:

On constate une déviation importante par rapport aux lois habituelles de l'arithmétique.

$x + (y + z)$ peut être différent de $(x + y) + z$.

Exemple :

Pour 4 chiffres significatifs ($s = 4$) on a :

$$(1 + 0.0005) + 0.0005 = 1.000$$

car

$$0.1 \times 10^1 + 0.5. \times 10^{-3} = 0.1. \times 10^1 + 0.00005. \times 10^1 = 0.1 \times 10^1 + 0.0000. \times 10^1 = 0.1 \times 10^1$$

et

$$1 + (0.0005 + 0.0005) = 1.001$$

Ainsi, l'addition flottante n'est pas associative .(TD:Sommation d'une série à termes positifs)

On constate aussi que si y est trés petit par rapport à x, l'addition de x et y donnera seulement x.

Exemple :

L'équation $1 + x = x$ a $x = 0$ comme unique solution. Mais dans un système à 10 chiffres significatifs, elle aura une infinité de solutions (il suffit de prendre $\mid x \mid \prec 5 \times 10^{-11}$)

La distributivité de la multiplication par rapport à l'addition.

Exemple :

Considérons l'opération

$$122 \times (333 + 695) = (122 \times 333) + (122 \times 695) = 125416$$

Si nous effectuons ces deux calculs en arithmétique à 3 chiffres ($s = 3$) et arrondi, nous obtenons:

$$\mathit{122} \times (\mathit{333} + \mathit{695}) = fl(\mathit{122}) \times fl(\mathit{1028})$$
$$= \mathit{122} \times \mathit{103} \times \mathit{10}^1 = fl(\mathit{125660}) = \mathit{126} \times \mathit{10}^3$$
$$(\mathit{122} \times \mathit{333}) + (\mathit{122} \times \mathit{695}) = fl(\mathit{40626}) + fl(\mathit{84790})$$
$$\mathit{406} \times \mathit{10}^2 + \mathit{848} \times \mathit{10}^2 = fl(\mathit{406} + \mathit{848}) \times \mathit{10}^2 = fl(\mathit{1254} \times \mathit{10}^2) = \mathit{125} \times \mathit{10}^3$$

Donc la distributivité de la multiplication par rapport à l'addition n'est pas respectée en arithmétique flottante.

0.3 Propagation des erreurs.

Une étude de la propagation des erreurs d'arrondi permattra d'expliquer ce phénomène.

Soit à calculer e^x à l'aide de son développement en série qui est convergent pour tout x :

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + ...$$

Il est evident que dans la pratique il est impossible d'effectuer la sommation d'une infinité de termes. On arrêtera donc lorsque le terme général $\frac{x^k}{k!}$ devient inférieur à 10^{-t}(on a t digits). Pour x négatif on sait que le reste de la serie est inférieur au premier terme négligé donc à 10^{-t}(puisque la serie est altérnée).

Les calculs suivant sont fait sur ordinateur pour $t = 14$.

x	e^x	S
-10	$4.54.10^{-5}$	$4.54.10^{-5}$
-15	$3.06.10^{-7}$	$3.05.10^{-7}$
-20	$2.06.10^{-9}$	$-1.55.10^{-7}$
-25	$1.39.10^{-11}$	$1.87.10^{-5}$
-30	$9.36.10^{-14}$	$6.25.10^{-4}$

On voit que pour $x \leq 20$ les résultats obtenus sont dépourvus de sens. L'explication de ce phénomène est la suivante: pour $x = -30$ les termes de la serie vont en croissant jusqu'à $\frac{x^{30}}{30!} = 8.10^{11}$ puis ils décroissent et $\frac{x^{107}}{107!} \tilde{} - 9.19.10^{-15}$.

L'erreur absolue sur le terme maximal est de $8.10^{11}.10^{-15} = 8.10^{-4}$. Ainsi le résultat obtenu pour S représente uniquement l'accumulation des erreurs d'arrondi sur les termes de plus grand module de développement en serie.

La propagation des erreurs est un des principaux problèmes en calcul numérique.

Considérons le cas d'une somme :

Dans l'addition, les erreurs absolues s'additionnent. Soit en effet ϵ_1 et $\epsilon_2$les erreurs absolues sur x_1 et x_2 .

On peut écrire:

$$(x_1 \pm \epsilon_1) + (x_2 \pm \epsilon_2) = (x_1 + x_2) \pm (\epsilon_1 + \epsilon_2)$$

En arithmétique flottante, l'erreur relative δ est à peu près constante et les erreurs absolues peuvent être approximativement explicités par :

$$\epsilon_1 = \mid x_1 \mid \times \delta, \epsilon_2 = \mid x_2 \mid \times \delta$$

Si les nombres en présence ont le même signe, l'erreur relative reste la même que celle qu'on avait x_1 et x_2. En effet

$$\frac{\epsilon_1 + \epsilon_2}{x_1 + x_2} = \frac{(\mid x_1 \mid + \mid x_2 \mid) \times \delta}{x_1 + x_2} = \pm\delta$$

Si par contre les nombres sont de signes différents, l'erreur relative peut être amplifiée de façon spectaculaire.

Dans la multiplication, les erreurs relatives s'additionnent.

En effet,soient x_1 et x_2 ,on a :

$$(x_1 + \epsilon_1) \times (x_2 + \epsilon_2) = x_1 x_2 + x_1 \epsilon_2 + x_2 \epsilon_1 + \epsilon_1 \epsilon_2$$

Si de plus les erreurs s'écrivent

$$\begin{aligned} \epsilon_1 &= |x_1| \times \delta \\ \epsilon_2 &= |x_2| \times \delta \end{aligned}$$

on a alors en négligeant certains termes:

$$\frac{|(x_1 + \epsilon_1) \times (x_2 + \epsilon_2) - x_1 x_2|}{x_1 x_2} = \frac{|x_1 \epsilon_2 + x_2 \epsilon_1|}{x_1 x_2} = \frac{\epsilon_1}{x_1} + \frac{\epsilon_2}{x_2} = 2\delta$$

Des formules équivalentes peuvent donner des résultats différents; on peut améliorer le résultant en utilisant une formule mathématique équivalente nécessitant des opérations différentes.

Exemple :
Si on considère les nombres $\sqrt{7001}$*et* .$\sqrt{7000}$.
En arirthmétique flottante à 8 chiffres , on a :

$$\sqrt{7001} = 0.83671979 \times 10^2$$
$$\sqrt{7000} = 0.83666003 \times 10^2$$

Donc

$$\sqrt{7001} - \sqrt{7000} = fl((0.83671979 - 0.83666003) \times 10^2) = 0.59760000 \times 10^{-2}$$

On peut obtenir un résultat plus précis en utilisant l'identité suivante:

$$\sqrt{x} - \sqrt{y} = (\sqrt{x} - \sqrt{y}) \times \frac{\sqrt{x} + \sqrt{y}}{\sqrt{x} + \sqrt{y}} = \frac{x - y}{\sqrt{x} + \sqrt{y}}$$

On obtient alors

$$\frac{1}{\sqrt{7001} + \sqrt{7000}} = \frac{1}{0.16733798 \times 10^3} = 0.59759297 \times 10^{-2}$$

0.4 Conditionnement et stabilité numérique.

Le fait que certains nombres ne soient pas représentés de façon exacte dans un ordinateur entraine que l'introduction même de donnée d'un problème en machine modifie quelque peu le problème initial; Il se peut que cette petite variation des données entraine une variation importante des résultats. C'est la notion de conditionnement d'un problème.

On dit qu'un problème est bien (ou mal) conditionné, si une petite variation des données entraine une petite (une grande) variation sur les résultats.

Cette notion de conditionnement est liée au problème mathématique lui même et est indépendante de la méthode utilisée pour le résoudre.

Une autre notion importante en pratique est celle de stabilité numérique. Un problème peut être bien conditionné et la méthode utilisée pour le résoudre peut être sujette à une propagation importante des erreurs numériques.

Ces notions de conditionnement d'un problème et de stabilité numérique d'une méthode de résolution sont fondamentales en analyse numérique. Si un problème est mal conditionné alors la solution exacte du problème tronqué ou arrondi à t digits pourra être très différente de la solution exacte du problème initial. Aucune methode ne pourra rien; il faudra essayer de donner une autre formulation au problème.

0.5 Instabilité numérique :

Si les erreurs introduites dans les étapes intermédiaires ont un effet négligeable sur le résultat final,on dira que le calcul ou l'algorithme est numériquement stable.si des petits changements sur les données entrainent des petits changements sur le résultat. Sinon , on dira que l'algorithme est numériquement instable.

Exemple :

On veut calculer la valeur de

$$I_n = \int_0^1 \frac{x^n}{a+x} dx$$

où a *est une constante plus grande que* 1*, pour plusieurs valeurs de* n*. pour ce faire, nous allons exprimer* I_n *récursivement, i.e. nous allons exprimer* I_n *en fonction de* n *et* I_{n-1}*.*

$$\begin{aligned}
I_n &= \int_0^1 \frac{x^{n-1}(x+a-a)}{a+x} dx \\
&= \int_0^1 x^{n-1} dx - a \int_0^1 \frac{x^{n-1}}{a+x} dx \\
&= \frac{1}{n} - aI_{n-1} \\
&= \sum_{i=0}^{n-1} \frac{(-a)^i}{n-i} + (-a)^n I_0
\end{aligned}$$

Comme

$$I_0 = ln(\frac{1+a}{a})$$

On peut calculer I_n *pour toutes les valeurs de* n.

Mais l'algorithme est numériquement instable car toute erreur dans le calcul de $I_0 = ln(\frac{1+a}{a})$ *va se propager.*

En effet si on note par I_0^* *la valeur approchée de* I_0 *et si* $I_0^* = I_0 + \epsilon$ *alors*

$$\begin{aligned} I_n^* &= \sum_{i=0}^{n-1} \frac{(-a)^i}{n-i} + (-a)^n \, I_0^* \\ &= \sum_{i=0}^{n-1} \frac{(-a)^i}{n-i} + (-a)^n \, (I_0 + \epsilon) \end{aligned}$$

donc $|I_n - I_n^*| \geq a^n \epsilon$.

Remarques:

Il y a en fait différentes sources d'erreur. Nous pouvons les classer en 3 catégories:

- *les erreurs liées à l'imprécision des mesures physiques ou au résultat d'un calcul approché*
- *les erreurs liées à l'algorithme utilisé*
- *les erreurs de calcul liées à la machine*

En général, pour l'objet de notre cours, si le premier chapitre met l'accent sur les erreurs liées à la machine, nous nous interresserons beaucoup plus aux erreurs liées aux méthodes ou encore aux algorithmes utilisés.

Série de Travaux Dirigés N I

Exercice 1 : En arithmétique flottante avec 3 chiffres significatifs et arrondi, illustrer la non-validité des lois d'associativité et de distributivité.

(On pourra prendre : $x = 854$, $y = 251$ et $z = 852$)

Exercice 2 : Soit p une fonction polynôme de degré n définie par

$$p(x) = \sum_{i=0}^{n} a_i x^i$$

Entrée:

$n, (a_i)_{0 \leq i \leq n}, t$

Sortie:

$val = p(t)$

Description du corps de l'algorithme:

$$a \longleftarrow a_n$$
$$\left\{ \begin{array}{l} pour\ i = n-1 \text{ à } 0 \text{ faire} \\ \quad a \longleftarrow a_i + t \times a \\ \qquad fin \end{array} \right.$$
$$val \longleftarrow a$$

Expliquer cet algorithme.

Exercice 3 : En arithmétique flottante, avec $s = 3$, calculer $\sum_{i=1}^{10} \frac{1}{i^2}$.

1) En calculant : $\frac{1}{1} + \frac{1}{4} + ... \frac{1}{100}$

2) En calculant : $\frac{1}{100} + \frac{1}{81} + ... + \frac{1}{1}$

Quel résultat est le plus précis et pourquoi?

Exercice 4 : Résolution d'équations du second degré

Soit l'équation du second degré avec c et $b \succ 0$

$$x^2 + bx + c = 0$$

On suppose que le discriminant $\Delta \succ 0$ est proche de b^2

1) Donner une expression de x_1 et $x_2 (x_1 \prec x_2)$

2) Montrer que la racine x est évaluée avec plus de précision en utilisant $x_1 = \dfrac{c}{x_2}$

3) Vérifier que pour $b = 160$ et $c = 1$, Δ est strictement positif et proche numériquement de b^2

Exercice 5 : On considère le polynôme $ax^2 + bx + c = 0$ $(a \neq 0)$. On suppose que le discriminant $\Delta \succ 0$. On sait que

$$(1) \qquad x_1 = \frac{-b+\sqrt{b^2-4ac}}{2a} \text{ et } x_2 = \frac{-b-\sqrt{b^2-4ac}}{2a}$$

1) Vérifier que $x_1 + x_2 = -\frac{b}{a}$ et $x_1 * x_2 = \frac{c}{a}$.

2) Utiliser ce résultat pour montrer que ces racines peuvent aussi s'écrire sous la forme

$$(2) \qquad x_1 = \frac{-2c}{b+\sqrt{b^2-4ac}} \text{ et } x_2 = \frac{-2c}{b-\sqrt{b^2-4ac}}$$

3) Pour $s = 4$ et trouver les racines de $x^2 + 53.1x + 1 = 0$ en calculant

i) x_1 à partir de (1) et la relation $x_1 * x_2 = \frac{c}{a}$.

ii) x_1 à partir de (2) et la relation $x_1 * x_2 = \frac{c}{a}$.

Quel calcul donne le meilleur résultat et pourquoi?

iii) Si on clacule d'abord x_2, laquelle des formules (1) et (2) serait-il préférable de choisir et pourquoi?

Exercice 6 : Résolvez ces deux systèmes linéaires:

$$(1) \qquad \begin{cases} x + y = 2 \\ x + 1.01y = 2.01 \end{cases}$$

$$(2) \qquad \begin{cases} x + y = 2 \\ x + 1.01y = 2.02 \end{cases}$$

Que remarquez-vous?

Exercice 7 : On cherche les racines de

$$p(x) = (x-1)(x-2)(x-10)$$

Elles sont évidentes.

Si on développe ce polynome en une valeur approchée pour l'une des racines par exemple 10.1

p(x) devient

$$p(x) = (x - 10.1)(x^2 + bx + c)$$

Calculer b et c .

En déduire les racines du polynomes du second degré. Que remarquez-vous?

Exercice 8:

Soient ϵ_1 et ϵ_2 les erreurs absolues sur ;x_1 et x_2 .

On peut écrire:

$$(x_1 \pm \epsilon_1) + (x_2 \pm \epsilon_2) = (x_1 + x_2) \pm (\epsilon_1 + \epsilon_2)$$

En arithmétique flottante, l'erreur relative δ est à peu près constante et les erreurs absolues peuvent être approximativement explicités par :

$\epsilon_1 =\mid x_1 \mid \times\delta, \epsilon_2 =\mid x_2 \mid \times\delta$

Que peut-on en conclure pour l'erreur relative obtenue pour la multiplication et l'addition?

Chapter 1

Résolution de f(x)=0

Introduction

Soit f une fonction numérique d'une variable réelle.

On cherche les racines simples de l'équation

$$f(x) = 0 \quad (1)$$

La première étape consiste à isoler les racines, c'est à dire trouver un intervalle $[a, b]$ dans lequel α est l'unique racine réelle de (1). On supposera que f est continue et dérivable autant de fois que nécessaire dans cet intervalle.

Pour trouver cet intervalle on aura besoin de quelques calculs préliminaires en utilisant soit le graphe des fonctions, soit (si la fonction f est continue dans $[a, b]$) le théorème des valeurs intermédiaires en calculant $f(a)$ et $f(b)$

Si $f(a) * f(b) \prec 0$ $\quad$ f admet un nombre impair de racines dans $[a, b]$

Si $f(a) * f(b) \succ 0$ $\quad$ f admet un nombre pair de racines

Exemple :

Soit. la fonction $f(x) = x - \frac{e^x}{e^x - 2}$

and Settings/ANO2005/Bureau/An-Num-deug1/graphics/ANum1-chap1-old

$_1.png$

$f(x) = x - \frac{e^x}{e^x - 2}$

La fonction n'est pas définie pour $x = ln(2)$ et on a $f'(x) = 1 + \frac{2e^x}{(e^x-2)^2}$

donc $f'(x) \succ 0$ pour tout x

L'équation a donc 2 racines simples situées de chaque côté de $ln(2)$.

On vérifie sans problème qu'une première racine appartient à $[-1, 0]$ et la deuxième à $[1, 2]$

On supposera donc désormais avoir trouvé un intervalle $[a, b]$ où f admet une unique racine simple et on supposera que f est définie, continue, et autant de fois continument dérivable que nécessaire.

Nous allons à présent définir la notion d'algorithme.

Nous appellerons algoritnme toute méthode de résolution d'un problème donné.

Pour tout problème, nous avons des données et des résultats. Les données sont appelées paramètres d'entrée (input) et les résultats paramètres de sortie (output). Ils constituent l'interface de l'algorithme (ou encore la partie visible de l'algorithme).

Dans ce chapitre, nous désignerons par$\{p_n\}$ une suite de nombres réels .

Il y a plusieurs façons de générer les termes d'une suite. En analyse numérique, on construit les suites à l'aide d'un procédé itératif appelé algorithme.

Les algorithmes classiques que nous allons étudier sont les suivants :

-Méthode de la bissection
-Méthode de Newton-Raphson
-Méthode de la sécante
-Méthode du point fixe

Méthode de la bissection

Considérons une fonction $f(x)$ quelconque, continue et cherchons p tel que $f(p) = 0$

Nous supposons qu'on a localisé par tatonnement un intervalle $[a, b]$ dans lequel la fonction change de signe (c.à.d. $f(a) * f(b) \prec 0$) on pose $c = \frac{a+b}{2}$, si $f(a) * f(c) \prec 0$ on remplace b par c sinon on remplace a par c, et on continue cette operation jusqu'à ce qu'on trouve p avec la précision demandée.

Algorithme de bissection (ou de dichotomie)

But : Donner une fonction continue $f(x)$ et un intervalle $[a, b]$ pour lequel $f(a)$ et $f(b)$ sont de signes contraires, trouver une solution de $f(x) = 0$ dans cet intervalle.

Entrées : a, b les extrémités de l'intervalle
ϵ la précision désirée
N_0 le nombre maximal d'itérations

Sortie : la valeur approchée de la solution de $f(p) = 0$

choix

Si $f(a) * f(b) > 0$: $\rightarrow'$pas de changement de signe$'$

Si $f(a) * f(b) \leq 0$: $n \longleftarrow 1$

iteration **arrêt-si** $n > N_0$ ou $\mid b - a \mid < \epsilon$ ou $f(a).f(b) = 0$

$p \longleftarrow \frac{a+b}{2}$
choix
Si $f(a) * f(p) \leq 0$: $b \longleftarrow p$
Si $f(a) * f(p) > 0$: $a \longleftarrow p$
fin-choix
$n \longleftarrow n+1$
fin-iteration
choix
Si $f(a)*f(b) = 0$: a ou b sont solutions
Si $\mid b - a \mid < \epsilon$: $\rightarrow a$ 'solution à ϵ près'
Sinon **:** 'le nombre maximum d'itération est atteit'
fin-choix
fin-choix

Cet algorithme peut aussi s'écrire sous la forme :

Etape 0: Si f(a)=0 imprimer la solution est a ,aller à l'étape 9
Si f(b)=0 imprimer la solution est b,aller à l'étape 9
Etape 1:
si f(b)f(a)
alors imprimer (il n'y a pas de changement de signe)
aller à l'étape 9
Etape 2:poser n=1
Etape 3:
Tant que N$\leq N_0$,faire les étapes 4 à 7
Etape 4:poser p=$\frac{a+b}{2}$
Etape 5:Si f(p)=0 ou $\frac{b-a}{2} \leq \epsilon$
Alors imprimer p
Fin
Etape 6: poser n=n+1
Etape 7 Si f(a)*f(p)$\succ$ 0
alors poser a=p
sinon poser b=p
Etape 8: Imprimer aprés N_0 itérations l'approximation obtenue est p et l'erreur maximale est $\frac{b-a}{2}$
Etape 9: Fin

Méthode de Newton-Raphson:
INSERER GRAPHE
But: Trouver une solution de $f(x) = 0$
Entrées: une approximation initiale p_0
ε(la précision désirée)
N_0 (le nombre maximum d'itérations)
Sortie: valeur approchée de p ou un message d'échec
Etape 3:poser$p = p_0 - \frac{f(p_0)}{f'(p_0)}$
Etape 4:Si$\mid p - p_0 \mid \leq \varepsilon$ alors imprimer p

Méthode de la sécante

La méthode de Newton -Raphson suppose le calcul de $f'(p)$ à chaque étape. Il se peut qu'on ne dispose pas d'un programme permettant de calculer systématiquement f' .

L'algorithme suivant peut être considéré comme une approximation de la méthode de Newton.

Au lieu d'utiliser la tangente au point p_n nous allons utiliser la sécante passant par les points d'abscisses p_n et p_{n-1} pour en déduire p_{n+1}.

L'équation de la sécante s'écrit :
$s(x)=f(p_n)+(x-p_n)\frac{f(p_n)-f(p_{n-1})}{p_n-p_{n-1}}$
Si $s(p_{n+1})=0$, on en déduit:
$p_{n+1}=p_n-f(p_n)\frac{p_n-p_{n-1}}{f(p_n)-f(p_{n-1})}$

INSERER GRAPHE
Algorithme de la sécante:

But: Trouver une solution de $f(x) = 0$
Entrées: deux approximations initiales p_0 et p_1
ε(la précision désirée)
N_0 (le nombre maximum d'itérations)
Sortie:la valeur approchée de p ou un message d'échec

Etape 1: poser N=2
$q_0=f(p_0)$
$q_1=f(p_1)$
Etape 2: Tant que $N \leq N_0 + 1$,faire les étapes 3 à 6
Etape 3:poser $p = p_1 - q_1\frac{(p_1-p_0)}{q_1-q_0}$

Etape 4:Si $| p - p_1 | \leq \varepsilon$ alors imprimer p

Fin

Etape 5: Poser N=N+1

Etape 6:Poser $p_0=p_1$

$q_0=q_1$

$p_1=p$

$q_1=f(p)$

Etape 7:Imprimer la méthode a échoué aprés N_0 itérations

Etape 8:Fin

Méthode du point fixe

Nous pouvons observer que la méthode de Newton peut s'interpréter comme $p_{n+1} = g(p_n)$ où

$g(x) = x - (\frac{f(x)}{f'(x)})$. Maintenant , si la fonction $g(x)$ est continue et si l'algorithme converge (c.à.d. $p_n \to p$),

on tire de $p_{n+1} = g(p_n)$ que p satisfait l'équation $p = g(p)$; on dit que p est un point fixe de g.

On peut toujours transformer un problème du type $f(x) = 0$ en un problème de la forme $x = g(x)$ et ce d'une infinité de façons.

Par exemple $x^2 - 2 = 0$ ou $x = 2/x$

ou $x = x^2 + x - 2$

ou $x = \alpha(x^2 - 2) + x$

Il faut toutefois noter que ce type de transformations introduisent des solutions 'parasites'.

Par exemple : résoudre $1/x = a$ ou encore $x = 2x - ax^2$

On voit que 0 est racine de la deuxième équation mais pas de la première.

Algorithme du point fixe

But: trouver une solution de $g(x) = x$

Entrées: une approximation initiale p_0

ε(la précision désirée)

N_0 le nombre maximale d'itérations

Sortie: valeur approchée de p ou un message d'échec

Etape 1: poser N=1

Etape 2 Tant que N$\leq N_0$,faire les étapes 3 à 6

Etape 3:poser p=g(p_0)

Etape 4 Si $\mid p - p_0 \mid \leq \varepsilon$

alors imprimer p

Fin

Etape 5:poser n=n+1

Etape 6:poser p_0=p

Etape 7: Imprimer (la méthode a échoué aprrès N_0 itérations)

Convergence des algorithmes

Ordre de convergence

Etude de la convergence des méthodes itératives à un pas

Ordre de convergence

Considérons une suite $\{p_n\}$ convergeanr vers p et posons $e_n = p_n - p$.

On dit dans le cas où $\left\{\left|\frac{e_n}{e_{n-1}}\right|\right\}$ converge, que la suite p_n converge linéairement vers p ou encore que la méthode est du premier ordre.

Si on a $\left\{\left|\frac{e_n}{(e_{n-1})^k}\right|\right\}$ converge, alors la convergence est dite d'ordre k

Par exemple la suite $p_n = \frac{1}{n}$est d'ordre 1

Série d'exercices n°2

Exercices 1

Résoudre à l'aide de la méthode de bisection $tanx - x = 0$ dans l'intervalle $[4; 4.7]$.

Exercice 2

On considère l'équation

(1) $e^x - 4x = 0$

1)Déterminer le nombre et la position approximative des racines de (1) situées dans .$x \geq 0$

2)Utiliser l'algorithme de bissection pour déterminer la plus petite de ces racines à ε près.(par exemple 10^{-7})

3)Sans faire d'itérations, déterminer combien vous devriez en faire pour calculer la plus grande racine à l'aide de la bissection avec une précision de 10^{-8}, si l'intervalle de départ est $[2; 2, 5]$

Exercice 3

Écrire un algorithme pour calculer par la méthode de Newton la racine K-ième d'un nombre.

Quelle est la valeur de $s = \sqrt{2 + \sqrt{2 + \sqrt{2 + ...}}}$?

Suggestion: écrire .p_{n+1}=G(p_n),p_0=0 Quel est le taux de convergence ?

Exercice 4

Écrire 3 méthodes itératives pour la résolution de $x^3 - x - 1$ =0 et vérifier expérimentalement leur convergence avec $x_0 = 1, 5$. Trouver à 10^{-6}près la racine comprise entre 1 et 2. Connaissant la valeur de cette racine, calculer le taux de convergence de vos 3 méthodes. Ce résultat coincide-t-il avec l'expérience?

Exercice 5

Résoudre x^2-1=0 en utilisant la méthode de la sécante avec $x_0 = -3$ et $x_1 = 5/3$. Qu'arrivera-t-il si on choisit et $x_0 = 5/3$ et $x_1 = -3$? Expliquez.

Université Mohammed V Agdal
Faculté des Sciences
Département de Mathématiques et Informatique
Avenue Ibn Batota, B.P. 1014
Rabat, Maroc

Filière :
Sciences Mathématiques et Informatique (SMI)
et
Sciences Mathématiques (SM)

Module Analyse Numérique I :

ANALYSE NUMERIQUE
Notes de Cours

Par

Pr. **Souad El Bernoussi** — bernous@fsr.ac.ma
Pr. **Saïd El Hajji** — elhajji@fsr.ac.ma
Pr. **Awatef Sayah**

htttp://www.fsr.ac.ma/ANO/

Année 2005-2006

TABLE DES MATIERES

Chapitre 1

Représentation des nombres en machine

1.1 Arithmétique des calculateurs et Sources d'erreurs

Si sophistiqué qu'il soit , un calculateur ne peut fournir que des réponses approximatives. Les approximations utilisées dépendent à la fois des contraintes physiques (espace mémoire, vitesse de l'horloge...) et du choix des méthodes retenues par le concepteur du programme . (pour plus de détails sur le fonctionnement d'un ordinateur et la terminologie de base voir par exemple la page web htttp://www.commentcamarche.com

Le but de ce chapitre est de prendre connaissance de l'impact de ces contraintes et de ces choix méthologiques. Dans certains cas il doit être pris en compte dans l'analyse des résultats dont une utilisation erronée pourrait être coûteuse.

La première contrainte est que le système numérique de l'ordinateur est discret, c'est à dire qu'il ne comporte qu'un nombre fini de nombres; Il en découle que tous les calculs sont entachés d'erreurs.

1.1.1 Evaluation de l'erreur

Rappelons d'abord quelques notion de base ;

Si X est une quantité à calculer et X^* la valeur calculée, on dit que :

1. $X - X^*$ est l'erreur et $\mid E \mid = \mid X - X^* \mid$ est l'erreur absolue.

Exemple :

Si $X = 2.224$ et $X^* = 2.223$ alors l'erreur absolue

$$| E |=| X - X^* |= 2.224 - 2.223 = 0.001$$

2. $E_r = \left|\frac{X-X^*}{X_r}\right|$ est l'erreur relative, $X_r \neq 0$. X_r est une valeur de référence pour X. En général ,on prend $X_r = X$.

   Exemple :

   Si $X = 2.224$ et $X^* = 2.223$ alors , si on prend $X_r = X$, l'erreur relative

   $$E_r = \left|\frac{X - X^*}{X_r}\right| = \frac{| X - X^* |}{| X |} = \frac{0.001}{2.224} = 4.496 \times 10^{-4}$$

   Cependant, si X est la valeur d'une fonction $F(t)$ avec $a \leq t \leq b$, on choisira parfois une valeur de référence globale pour toutes les valeurs de t.

   Exemple :

   SI $X = \sin(t)$ avec $0 \leq t \leq \frac{\pi}{4}$, on pourra prendre

   $$X = \frac{\sqrt{2}}{2} = \sup_{0 \leq t \leq \frac{\pi}{4}} \sin(t).$$

En général , on ne connait pas le signe de l'erreur de sorte que l'on considère les erreurs absolues et les erreurs relatives absolues.

Les opérations élémentaires propagent des erreurs.

Dans la pratique, on considère que :

1) L'erreur absolue sur une somme est la somme des erreurs absolues.

2) L'erreur relative sur un produit ou un quotient est la somme des ereurs relatives.

On peut estimer l'effet d'une erreur E sur l'argument x d'une fonction $f(x)$ au moyen de la dérivée de $f(x)$. En effet $f(x + E) \simeq f(x) + Ef'(x)$

Exemple :

Calculer la valeur de $(11111111)^2$

La valeur fournie par une petite calculatrice à cinq chiffres est $1,2345x10^{14}$

Mais la réponse exacte est 123456787654321.

La machine a donc tronqué le résultat à 5 chiffres et l'erreur absolue est de $6 * 19^9$.

L'erreur relative est de 0.0005% *.*

Cet exemple montre qu'il faut établir clairement l'objectif visé.

Cet objectif est double ;

1) Nous voulons un bon ordre de grandeur (ici 10^{14}) et avoir le maximum de décimales exactes,

2) Ce maximum ne peut excéder la longueur des mots permis par la machine et dépend donc de la machine

1.1.2 La mémoire de l'ordinateur : le stockage des nombres

La mémoire d'un ordinateur est formée d'un certain nombre d'unités adessables appelées OCTETS . Un ordinateur moderne contient des millions voir des milliards d'octets. Les nombres sont stockés dans un ordinateur comme ENTIERS ou REELS.

Les nombres entiers :

Les nombres entiers sont ceux que l'on utilise d'habitude sauf que le plus grand nombre représentable dépend du nombre d'octets utilisés:

-avec deux (2) octets, on peut représenter les entiers compris entre

$$-32768 \text{ et } 32767$$

-avec quatre (4) octets on peut représenterr les entiers compris entre

$$-2147483648 \text{ et } 2147483647$$

Les nombres réels

Dans la mémoire d'un ordinateur, les nombres réels sont représentés en notation flottante.

Cette notation a été introduite pour garder une erreur relative à peu prés constante; quelque soit l'ordre de gandeur du nombre qu'on manipule.

En notation flottante, un nombre a la forme:

$$x = \pm Y \times b^e$$

b est la base du système numérique utilisé

Y est la mantisse : une suite de s entier $y_1 y_2 ... y_s$ avec $y_1 \neq 0$ si $x \neq 0$ et $0 \leq y_i \leq (b-1)$

e est l'exposant(un nombre entier relatif)

La norme choisie est celle où la mantisse est comprise entre 0 et 1 et où le premier chiffre après la virgule est différent de zéro.

Calcul de l'erreur

Nous terminons ce chapitre en définissant les notions de troncature et d'arrondie.

Exemple :

En base 10, $x = 1/15 = 0.066666666......$

Dans le cas d'une représentation tronquée nous aurons, pour $s = 5$, $fl(x) = 0.66666 * 10^{-1}$.

Remarquez comment nous avons modifié l'exposant afin de respecter la règle qui veut que le premier chiffre de la mantisse ne soit pas nul .

Dans ce cas, l'erreur absolue $X - fl(X)$ est de $6 \times 10^{-7 \cdot}$. L'erreur relative est de l'ordre de 10^{-5}

Dans une représentation tronquée à s chiffres, l'erreur relative maximale est de l'ordre de 10^{-s}

Dans une représentation arrondie, lorsque la première décimale négligée est supérieure à 5, on ajoute 1 à la dernière décimale conservée.

Exemple :

$x = 1/15 = 0.066666666.$

Nous écrirons $fl(x) = 0.66667 \times 10^{-1}$

L'erreur absolue serait alors 3.333×10^{-7} *et l'erreur relative serait* 5×10^{-6}

En général, l'erreur relative dans une représentation arrondie à s chiffres est de $5 \times 10^{-(s+1)}$ soit la moitié de celle d'une représentation tronquée.

1.2 Les régles de base du modèle

Pour effectuer une opération sur deux nombres réels, on effectue l'opération sur leurs représentations flottantes et on prend ensuite la représentation flottante du résultat.

l'addition flottante

$$x \oplus y = fl(fl(x) + fl(y))$$

la soustraction flottante

$$x \ominus y = fl((x) - fl(y))$$

la multiplication flottante

$$x \otimes y = fl(fl(x) \times fl(y))$$

la division flottante

$$x \div y = fl(fl(x)/fl(y))$$

Chaque opération intermédiaire dans un calcul introduit une nouvelle erreur d'arrrondi ou de troncature.

Dans la pratique, il faudra se souvenir du fait que deux expressions algébriquement équivalentes peuvent fournir des résultats différents et que l'ordre des opérations peut changer les résultats.

Pour l'addition et la soustraction on ne peut effectuer ces 2 opérations que si les exposants sont les mêmes. On transforme le plus petit exposant et donc on ne respecte plus la régle voulant que le premier chiffre de la mantisse ne soit pas nul.

Quelques remarques sur ce modèle:

On constate une déviation importante par rapport aux lois habituelles de l'arithmétique.

$x + (y + z)$ peut être différent de $(x + y) + z$.

Exemple :

Pour 4 chiffres significatifs ($s = 4$) on a :

$$(1 + 0.0005) + 0.0005 = 1.000$$

car

$$0.1 \times 10^1 + 0.5. \times 10^{-3} = 0.1. \times 10^1 + 0.00005. \times 10^1 =$$
$$0.1 \times 10^1 + 0.0000. \times 10^1 = 0.1 \times 10^1$$

et

$$1 + (0.0005 + 0.0005) = 1.001$$

Ainsi, l'addition flottante n'est pas associative .(TD:Sommation d'une série à termes positifs)

On constate aussi que si y est trés petit par rapport à x, l'addition de x et y donnera seulement x.

Exemple :

L'équation $1 + x = x$ a $x = 0$ comme unique solution. Mais dans un système à 10 chiffres significatifs, elle aura une infinité de solutions (il suffit de prendre $\mid x \mid \prec 5 \times 10^{-11}$)

La distributivité de la multiplication par rapport à l'addition.

Exemple :

Considérons l'opération

$$122 \times (333 + 695) = (122 \times 333) + (122 \times 695) = 125416$$

Si nous effectuons ces deux calculs en arithmétique à 3 chiffres ($s = 3$) et arrondi, nous obtenons:

$$122 \times (333 + 695) = fl(122) \times fl(1028)$$
$$= 122 \times 103 \times 10^1 = fl(125660) = 126 \times 10^3$$
$$(122 \times 333) + (122 \times 695) = fl(40626) + fl(84790)$$
$$406 \times 10^2 + 848 \times 10^2 = fl(406 + 848) \times 10^2 = fl(1254 \times 10^2) = 125 \times 10^3$$

Donc la distributivité de la multiplication par rapport à l'addition n'est pas respectée en arithmétique flottante.

1.3 Propagation des erreurs.

Une étude de la propagation des erreurs d'arrondi permattra d'expliquer ce phénomène.

Soit à calculer e^x à l'aide de son développement en série qui est convergent pour tout x :

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + ...$$

Il est evident que dans la pratique il est impossible d'effectuer la sommation d'une infinité de termes. On arrêtera donc lorsque le terme général $\frac{x^k}{k!}$ devient inférieur à 10^{-t}(on a t digits). Pour x négatif on sait que le reste de la serie est inférieur au premier terme négligé donc à 10^{-t}(puisque la serie est altérnée).

Les calculs suivant sont fait sur ordinateur pour $t = 14$.

x	e^x	S
-10	$4.54.10^{-5}$	$4.54.10^{-5}$
-15	$3.06.10^{-7}$	$3.05.10^{-7}$
-20	$2.06.10^{-9}$	$-1.55.10^{-7}$
-25	$1.39.10^{-11}$	$1.87.10^{-5}$
-30	$9.36.10^{-14}$	$6.25.10^{-4}$

On voit que pour $x \leq 20$ les résultats obtenus sont dépourvus de sens. L'explication de ce phénomène est la suivante: pour $x = -30$ les termes de la serie vont en croissant jusqu'à $\frac{x^{30}}{30!} = 8.10^{11}$ puis ils décroissent et $\frac{x^{107}}{107!} \tilde{} - 9.19.10^{-15}$.

L'erreur absolue sur le terme maximal est de $8.10^{11}.10^{-15} = 8.10^{-4}$. Ainsi le résultat obtenu pour S représente uniquement l'accumulation des erreurs d'arrondi sur les termes de plus grand module de développement en serie.

1.3.1 Conditionnement et stabilité numérique.

Le fait que certains nombres ne soient pas représentés de façon exacte dans un ordinateur entraine que l'introduction même de donnée d'un problème en machine modifie quelque peu le problème initial; Il se peut que cette petite variation des données entraine une variation importante des résultats. C'est la notion de conditionnement d'un problème.

On dit qu'un problème est bien (ou mal) conditionné, si une petite variation des données entraine une petite (une grande) variation sur les résultats.

Cette notion de conditionnement est liée au problème mathématique lui même et est indépendante de la méthode utilisée pour le résoudre.

Une autre notion importante en pratique est celle de stabilité numérique. Un problème peut être bien conditionné et la méthode utilisée pour le résoudre peut être sujette à une propagation importante des erreurs numériques.

Ces notions de conditionnement d'un problème et de stabilité numérique d'une méthode de résolution sont fondamentales en analyse numérique. Si un problème est mal conditionné alors la solution exacte du problèmetronqué ou arrondi à t digits pourra être très différente de la solution exacte du problème initial. Aucune methode ne pourra rien; il faudra essayer de donner une autre formulation au problème.

[1]

[1]S. El Bernoussi, S. El Hajji et A. Sayah

Chapitre 2

Résolution de f(x)=0

2.1 Introduction

Soit f une fonction numérique d'une variable réelle.

On cherche les racines simples de l'équation

$$(1) \qquad f(x) = 0$$

La première étape consiste à isoler les racines, c'est à dire trouver un intervalle $[a, b]$ dans lequel α est l'unique racine réelle de (1). On supposera que f est continue et dérivable autant de fois que nécessaire dans cet intervalle.

Pour trouver cet intervalle on aura besoin de quelques calculs préliminaires en utilisant soit le graphe des fonctions, soit (si la fonction f est continue dans $[a, b]$) le théorème des valeurs intermédiaires en calculant $f(a)$ et $f(b)$

Si $f(a) * f(b) < 0$ $\quad f$ admet un nombre impair de racines dans $[a, b]$

Si $f(a) * f(b) > 0$ $\quad f$ admet un nombre pair de racines

Exemple :

Soit. la fonction du grahe suivant :

La fonction n'est pas définie pour $x = ln(2)$ et on a $f'(x) = 1 + \frac{2e^x}{(e^x-2)^2}$

donc $f'(x) > 0$ pour tout x.

L'équation a donc 2 racines simples situées de chaque côté de $ln(2)$.

On vérifie sans problème qu'une première racine appartient à $[-1, 0]$ et la deuxième à $[1, 2]$

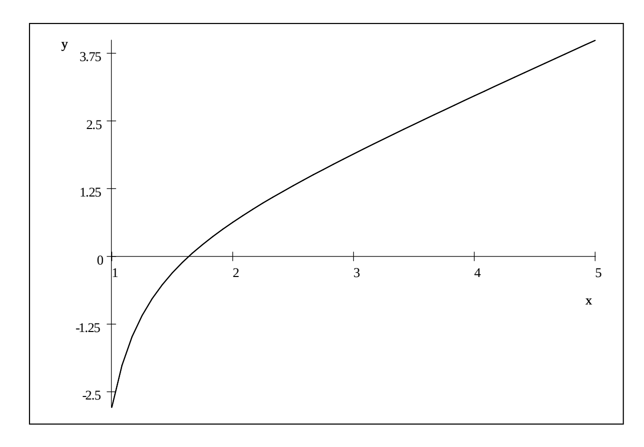


Figure 2.1: $f(x) = x - \frac{e^x}{e^x - 2}$

On supposera donc désormais avoir trouvé un intervalle $[a, b]$ où f admet une unique racine simple et on supposera que f est définie, continue, et autant de fois continument dérivable que nécessaire.

Nous allons à présent définir la notion d'algorithme.

Définition : Nous appellerons algoritnme toute méthode de résolution d'un problème donné.

Pour tout problème, nous avons des données et des résultats. Les données sont appelées paramètres d'entrée (input) et les résultats paramètres de sortie (output). Ils constituent l'interface de l'algorithme (ou encore la partie visible de l'algorithme).

Dans ce chapitre, nous désignerons par $\{p_n\}$ une suite de nombres réels .

Il y a plusieurs façons de générer les termes d'une suite. En analyse numérique, on construit les suites à l'aide d'un procédé itératif appelé algorithme.

Les algorithmes classiques que nous allons étudier sont les suivants:

i) Méthode de la bissection
ii) Méthode de Newton-Raphson
iii) Méthode de la sécante
iv) Méthode du point fixe.

Le but de ce chapitre est de trouver des approximations de la solution de l'équation (1) avec une précision donnée et un nombre d'itérations maximum.

Afin de comparer ces différentes méthodes, nous allons introduire la notion d'ordre de convergence.

2.2 Méthode de la bissection.

Considérons une fonction $f(x)$ quelconque, continue et cherchons p tel que $f(p) = 0$.

Nous supposons qu'on a localisé par tatonnement un intervalle $[a, b]$ dans lequel la fonction change de signe (c.à.d. $f(a) * f(b) \prec 0$) on pose $c = \frac{a+b}{2}$, si $f(a) * f(c) \prec 0$ on remplace b par c sinon on remplace a par c, et on continue cette operation jusqu'à ce qu'on trouve p avec la précision demandée.

Algorithme de bissection (ou de dichotomie)

But : Donner une fonction continue $f(x)$ et un intervalle $[a, b]$ pour lequel $f(a)$ et $f(b)$ sont de signes contraires, trouver une approximation de la solution de $f(x) = 0$ dans cet intervalle; en construisant une suite d'intervalles $([a_n, b_n])_n$ contenant cette racine et tets que a_n ou b_n est le milieu de l'intervalle $[a_{n-1}, b_{n-1}]$.

Entrées : a, b les extrémités de l'intervalle
ϵ la précision désirée
N_0 le nombre maximal d'itérations

Sortie : la valeur approchée de la solution de $f(p) = 0$

Etape 0: Si $f(a) = 0$ imprimer la solution est a, aller à l'étape 9
Si $f(b) = 0$ imprimer la solution est b, aller à l'étape 9

Etape 1:
si $f(b) * f(a) > 0$
alors imprimer (il n'y a pas de changement de signe)
aller à l'étape 9

Etape 2: poser $N = 1$

Etape 3:
Tant que $N \leq N_0$,faire les étapes 4 à 7
Etape4: poser $p = \frac{a+b}{2}$
Etape 5: Si $f(p) = 0$ ou $\frac{b-a}{2} \leq \epsilon$
Alors imprimer p
aller à l'étape 9
Etape 6: poser $N = N + 1$

Etape 7: Si $f(a) * f(p) > 0$
alors poser $a = p$
sinon poser $b = p$

Etape 8: Imprimer aprés N_0 itérations l'approximation obtenue est p et l'erreur maximale est $\frac{b-a}{2}$

Etape 9: Fin

2.3 Méthode de Newton-Raphson:

Le principe consiste à construire une suite $(x_n)_n$, telle que x_{n+1} soit l'intersection de la tangente à la courbe de f au point $(x_n, f(x_n))$ avec l'axe horizontal.

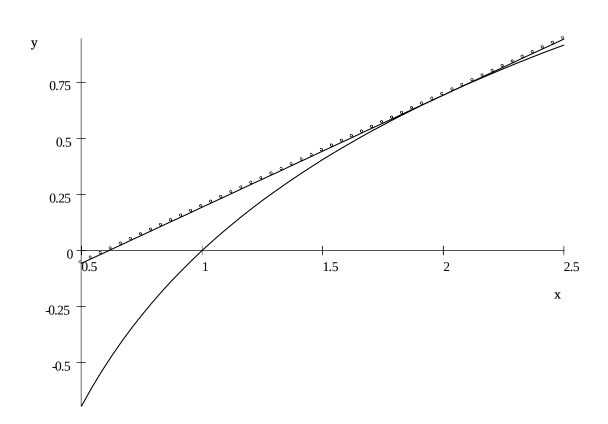


Figure 2.2: Méthode de Newton pour $f(x) = \log(x), x_0 = 2$.

On a:

$$\begin{cases} A = (x_0, f(x_0)), B = (x_1, 0) \in axe(Ox) \\ A \text{ et } B \in D : y = ax + b \end{cases}$$

donc

$$\begin{cases} f(x_0) = ax_0 + b \\ 0 = ax_1 + b \end{cases} \Rightarrow \begin{cases} a = f'(x_0) \\ x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} \end{cases}$$

Algorithme de Newton-Raphson.

But: Trouver une solution de $f(x) = 0$

Entrées: une approximation initiale p_0
ε (la précision désirée)
N_0 (le nombre maximum d'itérations)
Sortie: valeur approchée de p ou un message d'échec
Etape1 : $N = 1$
Etape 2: Tant que $N \leq N_0$, faire les étapes 3 à 6.
Etape 3: Poser $p = p_0 - \frac{f(p_0)}{f'(p_0)}$
Etape 4: Si$\mid p - p_0 \mid \leq \varepsilon$ alors imprimer p
aller à l'étape 8.
Etape 5: Poser $N = N + 1$.
Etape 6: Poser $p_0 = p$.
Etape 7: Imprimer la méthode a échoué après N itérations.
Etape 8: Fin.

2.4 Méthode de la sécante

La méthode de Newton-Raphson suppose le calcul de $f'(p)$ à chaque étape. Il se peut qu'on ne dispose pas d'un programme permettant de calculer systématiquement f' .

L'algorithme suivant peut être considéré comme une approximation de la méthode de Newton.

Le principe consiste à construire une suite $(x_n)_n$ à l'aide de la formule obtenue en remplaçant dans la méthode de Newton $f'(p_n)$ par $\frac{f(p_n)-f(p_{n-1})}{p_n-p_{n-1}}$. Ainsi au lieu d'utiliser la tangente au point p_n nous allons utiliser la sécante passant par les points d'abscisses p_n et p_{n-1} pour en déduire p_{n+1}. Ce dernier est obtenu comme intersection de la sécante passant par les points d'abscisse p_n et p_{n-1} et de l'axe des abscisses.

L'équation de la sécante s'écrit :
$s(x) = f(p_n) + (x - p_n)\frac{f(p_n)-f(p_{n-1})}{p_n-p_{n-1}}$
Si $s(p_{n+1}) = 0$, on en déduit:
$p_{n+1} = p_n - f(p_n)\frac{p_n-p_{n-1}}{f(p_n)-f(p_{n-1})}$

Algorithme de la sécante:

But: Trouver une solution de $f(x) = 0$
Entrées: deux approximations initiales p_0 et p_1

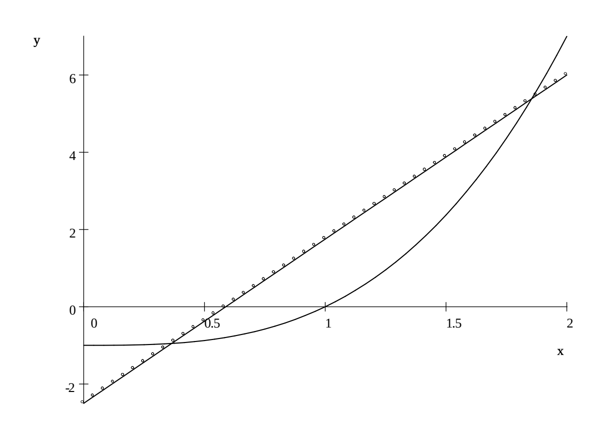


Figure 2.3: $f(x) = x^3 - 1$

ε (la précision désirée)
N_0 (le nombre maximum d'itérations)

Sortie:la valeur approchée de p ou un message d'échec

Etape 1: poser $N = 1$
$q_0 = f(p_0)$
$q_1 = f(p_1)$

Etape 2: Tant que $N \leq N_0 + 1$,faire les étapes 3 à 6

Etape 3: poser $p = p_1 - q_1\frac{(p_1 - p_0)}{q_1 - q_0}$

Etape 4: Si $\mid p - p_1 \mid \leq \varepsilon$ alors imprimer p
aller à l'étape 8

Etape 5: Poser $N = N + 1$

Etape 6: Poser $p_0 = p_1$
$q_0 = q_1$
$p_1 = p$
$q_1 = f(p)$

Etape 7: Imprimer la méthode a échoué aprés N_0 itérations

Etape 8: Fin.

2.5 Méthode du point fixe

Nous pouvons observer que la méthode de Newton peut s'interpréter comme $p_{n+1} = g(p_n)$ où

$g(x) = x - (\frac{f(x)}{f'(x)})$. Maintenant , si la fonction $g(x)$ est continue et si l'algorithme converge (c.à.d. $p_n \to p$), on tire de $p_{n+1} = g(p_n)$ que p satisfait l'équation $p = g(p)$; on dit que p est un point fixe de g.

On peut toujours transformer un problème du type $f(x) = 0$ en un problème de la forme $x = g(x)$ et ce d'une infinité de façons.

Par exemple

$x^2 - 2 = 0$

ou $x = 2/x$

ou $x = x^2 + x - 2$

ou $x = \alpha(x^2 - 2) + x$

Il faut toutefois noter que ce type de transformations introduisent des solutions 'parasites'.

Par exemple : résoudre $1/x = a$ ou encore $x = 2x - ax^2$

On voit que 0 est racine de la deuxième équation mais pas de la première.

Algorithme du point fixe

But: trouver une solution de $g(x) = x$

Entrées: une approximation initiale p_0

ε(la précision désirée)

N_0 le nombre maximale d'itérations

Sortie: valeur approchée de p ou un message d'échec

Etape 1: poser $N = 1$

Etape 2: Tant que $N \leq N_0$,faire les étapes 3 à 6

Etape 3: poser $p = g(p_0)$

Etape 4: Si $\mid p - p_0 \mid \leq \varepsilon$

alors imprimer p

aller à l'étape 8

Etape 5: poser $n = n + 1$

Etape 6: poser $p_0 = p$

Etape 7: Imprimer (la méthode a échoué aprrès N_0 itérations)

Etape 8 : Fin.

2.6 Convergence et ordre de convergence.

Définition: Soit D une partie de $\mathbb{R}$ et F une application de D dans D. On dit que la fonction F est contractante si

$$\forall x, y \in D \ , \exists k \in [0,1[\quad \text{tel que}$$
$$\mid F(x) - F(y) \mid \leq k \ \mid x - y \mid .$$

k est le coéfficient de contraction ou de Lipschitz de F.

Théorème: Considérons le segment $S = [p_0 - a, p_0 + a] \subset D$; si F est contractante sur S et si $\mid F(p_0) - p_0 \mid \leq (1-k)\, a$, alors l'itération $p_{n+1} = F(p_n)$ de point initial p_0 , converge vers l'unique point fixe $p \in S$ de F.

Théorème: Convergence locale.

Si F est différentiable au voisinage d'un point fixe p et si $\mid F'(p) \mid < 1$ alors :

$\exists V$ voisinage de p tels que $p_0 \in V$ et $p_{n+1} = F(p_n)$ converge vers p.

2.6.1 Interprétation graphique.

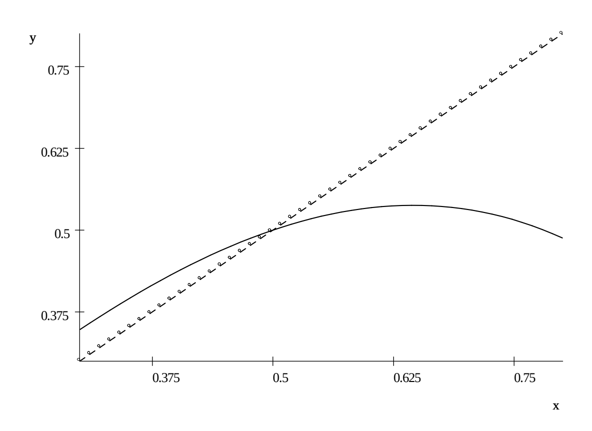
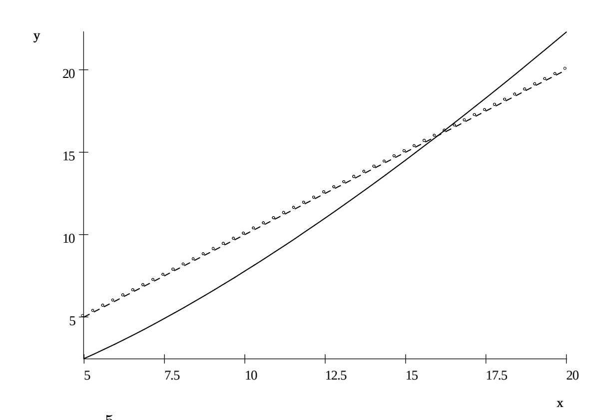


Figure 2.4: $F(x) = -x^3 + \frac{5}{4}x$; $|F'(x)| < 1$, convergence.

On voit graphiquement que $\mid F'(p) \mid < 1$, et par conséquent les itérations convergent vers le point fixe. p est un point fixe attractif. Par contre si $\mid F'(p) \mid > 1$ pas de convergence vers le point fixe, p est un point fixe répulsif.



$F(x) = x^{\frac{5}{4}} - x, |F'(16)| > 1$, l'itération diverge.

Remarque: Un point fixe répulsif pour une méthode devient attractif pour une autre.

2.6.2 Ordre de convergence.

La convergence de l'itération $p_{n+1} = F(p_n)$ vers le point fixe peut se faire plus ou moins vite.

Définition : Considérons une suite $\{p_n\}$ convergeant vers p et posons $e_n = p_n - p$.

On dit dans le cas où $\left\{\left|\frac{e_n}{e_{n-1}}\right|\right\}$ converge, que la suite p_n converge linéairement vers p ou encore que la méthode est du premier ordre.

Si on a $\left\{\left|\frac{e_n}{(e_{n-1})^k}\right|\right\}$ converge, alors la convergence est dite d'ordre k.

Exemple :

La méthode de Newton pour résoudre l'équation $f(x) = 0$ est une méthode de type point fixe avec $F(x) = x - \frac{f(x)}{f'(x)}$. Si x^* est racine simple de $f(x) = 0$, alors $f'(x^*) \neq 0$ et il existe un voisinage V de x^* tel que pour tout $p_0 \in V$, la suite $(p_n)_n$ converge vers x^* et l'ordre de convergence est 2.

(en effet $F'(x) = 1 - \frac{(f'(x))^2 - f(x)*f''(x)}{(f'(x))^2} \Rightarrow F'(x^*) = 0$. Ainsi d'aprés le théorème précédent la méthode de Newton converge. Pour déterminer l'ordre de convergence on utilise la formule de Tylors en x^* : $F(x) = F(x^*) + F'(x^*)(x - x^*) + F''(\theta x)\frac{(x-x^*)^2}{2}$).

2.7 Exercices

Série $f(x) = 0$

Exercices 1

Résoudre à l'aide de la méthode de bisection $tanx - x = 0$ dans l'intervalle $[4; 4.7]$.

Exercice 2

On considère l'équation

(1) $e^x - 4x = 0$

1) Déterminer le nombre et la position approximative des racines de (1) situées dans .$x \geq 0$

2) Utiliser l'algorithme de bissection pour déterminer la plus petite de ces racines à ε près.(par exemple 10^{-7})

3) Sans faire d'itérations, déterminer combien vous devriez en faire pour calculer la plus grande racine à l'aide de la bissection avec une précision de 10^{-8}, si l'intervalle de départ est $[2; 2, 5]$

Exercice 3

Écrire un algorithme pour calculer par la méthode de Newton la racine K-ième d'un nombre.

Quelle est la valeur de $s = \sqrt{2 + \sqrt{2 + \sqrt{2 + ...}}}$?

Suggestion: écrire .p_{n+1}=G(p_n),p_0=0 Quel est l'ordre de convergence ?

Exercice 4

Écrire 3 méthodes itératives pour la résolution de $x^3 - x - 1$ =0 et vérifier expérimentalement leur convergence avec $x_0 = 1, 5$. Trouver à 10^{-6}près la racine comprise entre 1 et 2. Connaissant la valeur de cette racine, calculer l'ordre de convergence de vos 3 méthodes. Ce résultat coincide-t-il avec l'expérience?

Exercice 5

Résoudre x^2-1=0 en utilisant la méthode de la sécante avec $x_0 = -3$ et $x_1 = 5/3$. Qu'arrivera-t-il si on choisit et $x_0 = 5/3$ et $x_1 = -3$? Expliquez. [1]

---

[1] S. El Bernoussi, S. El Hajji et A. Sayah

Chapitre 3

Algèbre linéaire

3.1 Introduction

Un sytème linèaire s'écrit sous la forme :

$$(1) \qquad Ax = b$$

où A est une matrice nxn à coefficients réels, $b \in \mathbb{R}^n$ et $x \in \mathbb{R}^n$.

La résolution de grands systèmes linéaires (et non linéaires) est pratique courante de nos jours. Elle apparait dans tous les domaines où l'on s'intéresse à la résolution numérique d'équations aux dérivées partielles.

Il existe plusieurs packages (linpack, eispack, ..), logiciels (Maple et Matlab) et programmes (http://www.netlib.com, numerical recipes, NAG, IMSL, ...) de base pour le résoudre.

Le choix de la méthode dépend fortement du type (forme) de la matrice.

Les méthodes de résolution sont de deux types :

Les méthodes directes : Une méthode est dite directe si elle permet d'obtenir la solution en un nombre fini d'opérations.

Les méthodes itératives : Une méthode est dite itérative si elle permet de construire une suite $(x_n)_n$ qui converge vers la solution.

Dans ce chapitre nous allons :

1. Rapeler des notions et notations de base relatives aux systèmes linéaires et aux matrices

2. Etudier une méthode directe : la méthode de Gauss.

3. Etudier la décomposition (factorisation) LU.

4. Etudier des applications : Inverse de matrices,...

3.2 Rappels sur les systémes linéaires

Un système de n équations linéaires à n inconnues peut toujours s'écrire sous la forme :

$$(1) \qquad Ax = b$$

où A est une matrice (a_{ij}) et x et b sont des vecteurs colonnes de dimension n.

Si la matrice A est inversible alors le système linéaire (1) admet une unique solution $x = A^{-1}b$ où A^{-1} est la matrice inverse de A.

Ainsi théoriquement le problème revient à calculer A^{-1}? Mais en pratique ce calcul est difficile.

Il existe plusieurs méthodes classiques pour résoudre (1) sans calculer A^{-1}.

Pour cela on va considerer le cas simple suivant :

$$(1) \qquad \begin{cases} x + 2y = 5 \\ 2x + y = 4 \end{cases}$$

i) La méthode de Cramer consiste à calculer la solution en calculant des déterminants.

On a: $x = \dfrac{\begin{vmatrix} 5 & 2 \\ 4 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix}} = \frac{-3}{-3} = 1$ et $y = \dfrac{\begin{vmatrix} 1 & 5 \\ 2 & 4 \end{vmatrix}}{\begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix}} = \frac{-6}{-3} = 2$

ii) La méthode de substitution (ou d'élimination) consiste à transformer le systéme (1).

$$(1) \qquad \begin{cases} x + 2y = 5 \\ 2x + y = 4 \end{cases} \Rightarrow \begin{cases} x = -2y + 5 \\ 2x + y = 4 \end{cases} \Rightarrow \begin{cases} x = -2y + 5 \\ 2(-2y + 5) + y = 4 \end{cases}$$

$$\Rightarrow \begin{cases} x = -2y+5 \\ 3y = 6 \end{cases} \Rightarrow \begin{cases} x = -2y+5 \\ y = 2 \end{cases} \Rightarrow \begin{cases} x = 1 \\ y = 2 \end{cases}$$

Peut-on généraliser ces méthodes pour un systéme de n équations avec $n \in \mathbb{N}$?

Théoriquement OUI mais en pratique cela va nécessiter beaucoup de calculs et de techniques.

3.3 Méthode Gauss

La méthode de résolution la plus étudiée (et une des plus employées) s'appelle méthode d'élimination de Gauss.

L'idée de base de cette méthode consiste à transformer le systéme linéaire (1) en un problème que l'on sait résoudre.

Si la matrice $A = D$ avec D une matrice diagonale, alors on sait résoudre (1).

Mais toute matrice n'est pas diagonalisable.

Si la matrice $A = U$ (ou L) avec U (ou T) une matrice triangulaire supérieure (ou inférieure) alors on sait résoudre (1).

Problème : Comment tranformer une matrice en une matrice triangulaire inférieure ou supérieure ?

La méthode de substitution (d'élimination) répond à cette question mais elle n'est pas automatique.

La méthode d'élimination de Gauss a pour but de remplacer le système (1) par un système triangulaire possédant la même solution. Son principe s'apparente à celui de la méthode de substitution (d'élimination) mais (comme on le verra ci dessous), il est plus simple à automatiser.

Regardons son fonctionnement sur l'exemple suivant cas $n = 3$:

On pose $A = (a_{ij})_{i,j=1,3}$ $X = (x_i)_{i=1,3}$et $b = (b_i)_{i=1,3}$ de telle sorte que $AX = b$ s'écrit sous la forme :

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3 \end{cases}$$

ou encore sous la forme dite augmentée

$$(A\ b) = \begin{pmatrix} a_{11} & a_{12} & a_{13} & b_1 \\ a_{21} & a_{22} & a_{23} & b_2 \\ a_{31} & a_{32} & a_{33} & b_3 \end{pmatrix}$$

On suppose que $a_{11} \neq 0$, par élimination, on obtient :

$$(A_1 \ b_1) = \begin{pmatrix} a_{11} & a_{12} & a_{13} & b_1 \\ 0 & a'_{22} & a'_{23} & b'_2 \\ 0 & a'_{32} & a'_{33} & b'_3 \end{pmatrix}$$

On va illustrer la méthode de Gauss sans passer par le sytéme augmenter :

On a :

$$(1) \left\{ \begin{array}{ll} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1 & (l1) \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2 & (l_2) \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3 & (l_3) \end{array} \right.$$

On note par (l_i) la $i^{ème}$ équation du système précedent.
On suppose que $a_{11} \neq 0$,
On pose :
$(l'_2) = a_{11}(l_2) - a_{21}(l'_1)$
et
$(l'_3) = a_{11}(l_3) - a_{31}(l'_1)$
Alors (1) s'écrit

$$(2) \left\{ \begin{array}{rl} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1 & (l1) \\ a'_{22}x_2 + a'_{23}x_3 = b'_2 & (l'_2) \\ a'_{32}x_2 + a'_{33}x_3 = b'_3 & (l'_3) \end{array} \right.$$

On suppose que $a'_{22} \neq 0$,
On pose :
$(l''_3) = a'_{22}(l'_3) - a'_{32}(l''_2)$
Alors (2) s'écrit

$$(2) \left\{ \begin{array}{rl} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1 & (l1) \\ a'_{22}x_2 + a'_{23}x_3 = b'_2 & (l'_2) \\ a^{''}_{33}x_3 = b^{''}_3 & (l^{''}_3) \end{array} \right.$$

Remarque :
i) Les termes diagonaux à chaque étape sont appelés les pivots,
ii) Si un pivot a_{ii} est nul on change de ligne (on permute) de i à n (pivotage partiel)

iii) Cette méthode se généralise assez facilement bien qu'il faut être prudent avec le choix du pivot. En pratique, il faut éviter de prendre des pivots "trop" petits.

Exemple : Sur l'importance du pivot

1) On considére le systéme :

$$(1) \qquad \begin{cases} x + y = 2 \\ 10^{-4}x + y = 1 \end{cases}$$

Calculer la solution exacte de ce systéme.

2) Calculer la solution pour $s = 3$ avec troncature des systèmes

$$(1) \qquad \begin{cases} x + y = 2 \\ 10^{-4}x + y = 1 \end{cases} \qquad et \qquad (1) \qquad \begin{cases} 10^{-4}x + y = 1 \\ x + y = 2 \end{cases}$$

Remarque : L'algorithme de Gauss est une méthode systématique de résolution de systèmes d'équations comportant un nombre quelconque d'inconnues.

Dans le cas où tous les pivots sont non nuls i.e. $a_{ii} \neq 0$, l'algorithme: Élimination de Gauss s'ecrit :

Partie 1: Réduction à la forme triangulaire (ou élimination de Gauss)

Entrée A et b
Sortie $A = U$ (forme triangulaire), et b.
Pour $j = 1, ..., (n-1)$
    Pour $i = j+1, ..., n$
        $l_{ij} \leftarrow \frac{a_{ij}}{a_{jj}}$
        Pour $k = j+1, ..., n$
            $a_{ik} \leftarrow a_{ik} - l_{ij}a_{jk}$
        Fin
        $b_j \leftarrow b_j - l_{ij}b_j$
    Fin
Fin
Sortie $A = U$ (forme triangulaire), et b

Cette partie s'écrit sous la forme (algorithmique)
Étape 1 : Poser $j = 1$
Étape 2: Tant que $j \leq n-1$ faire
    Étape 3: Si $a_{jj} = 0$ afficher 'pivot nul' aller à étape 14, sinon

Étape 4: Poser $i = j + 1$
Étape 5: Tant que $i \leq n$ faire
Étape 6: $l_{ij} = \frac{a_{ij}}{a_{jj}}$
Étape 7: Si $l_{ij} = 0$; aller à l'étape 12.
Étape 8: Poser $k = j + 1$
Étape 9: Tant que $k \leq n$ faire
Étape 10: $a_{ik} = a_{ik} - l_{ij}a_{jk}$, $k = k + 1$; aller à l'étape 9.
Étape 11: $b_i = b_i - l_{ij}b_j$;
Étape 12: poser $i = i + 1$; Aller à l'étape 5.
Étape 13: $j = j + 1$; Aller à l'étape 2.
Étape 14: Fin.

Remarques: Les éléments sous la diagonale principale de la nouvelle matrice obtenue sont nuls. Comme ils n'interviennent pas dans la résolution du système triangulaire formé, il est inutile que l'algorithme leur assigne cette valeur nulle.

Exemple : On considère le système linéaire :

$$\begin{cases} x + y + 3t = 4 \\ 2x + y - z + t = 1 \\ 3x - y - z + 2t = -3 \\ -x + 2y + 3z - t = 4 \end{cases}$$

qui s'écrit encore:

$$\begin{pmatrix} 1 & 1 & 0 & 3 \\ 2 & 1 & -1 & 1 \\ 3 & -1 & -1 & 2 \\ -1 & 2 & 3 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \\ -3 \\ 4 \end{pmatrix}$$

Nous appliquons l'algorithme à notre exemple en travaillant sur la matrice augmentée.

Nous obtenons

$$\left[\begin{array}{ccccc} & & A & & & b \\ 1 & 1 & 0 & 3 & . & 4 \\ 2 & 1 & -1 & 1 & . & 1 \\ 3 & -1 & -1 & 2 & . & -3 \\ -1 & 2 & 3 & -1 & . & 4 \end{array}\right]$$

$$\begin{bmatrix} 1 & 1 & 0 & 3 & . & 4 \\ 0 & -1 & -1 & -5 & . & -7 \\ 0 & -4 & -1 & -7 & . & -15 \\ 0 & 3 & 3 & 2 & . & 8 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & 0 & 3 & . & 4 \\ 0 & -1 & -1 & -5 & . & -7 \\ 0 & 0 & 3 & 13 & . & 13 \\ 0 & 0 & 0 & -13 & . & -13 \end{bmatrix}$$

Que l'on peut écrire sous la forme :

$$\left\{ \begin{array}{c} x + y + 3t = 4 \\ -y - z - 5t = -7 \\ 3z + 13t = 13 \\ -13t = -13 \end{array} \right. ,$$

Notons que l'étape $j = 3$ nous donnerait $l_{43} = 0$.
Nous avons maintenant un système triangulaire à résoudre.

Partie 2 : Remontée triangulaire

Entrée A, b avec A matrice triangulaire supérieure
Sortie x solution du sytème $Ax = b$

- Étape 1: $x_n = \frac{b_n}{a_{nn}}$
- Étape 2: Pour $i = n - 1, n - 2, ..., 1$ faire:

$$x_i = \frac{1}{a_{ii}}(b_i - \sum_{j=i+1}^{n} a_{ij}x_j)$$

En appliquant cet algorithme à notre exemple, nous obtenons $x = (-1, 2, 0, 1)$.

Remarque:

1. Dans la pratique le test (3) de l'algorithme d'élimination de Gauss ne conduit pas à l'arrêt. En fait, si le pivot est nul, on cherche, dans la même colonne, un élément d'indice plus grand non nul, puis on échange les lignes correspondantes. Si ceci est impossible, le système est singulier.

2. On est parfois amené, pour des raisons de stabilité numérique, à effectuer des échanges de lignes même si le test (3) est négatif (c'est à dire que le pivot est non nul). Ceci conduit à des stratégies dites de pivot que nous n'étudierons pas ici.

   Exemple : Résolution du système suivant :

$$\begin{cases} 2x+6y+10z=0 \\ x+3y+3z=2 \\ 3x+14y+28z=-8 \end{cases} \Leftrightarrow \begin{pmatrix} 2 & 6 & 10 \\ 1 & 3 & 3 \\ 3 & 14 & 28 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \\ -8 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 6 & 10 & 0 \\ 1 & 3 & 3 & 2 \\ 3 & 14 & 28 & -8 \end{pmatrix} \Rightarrow \begin{pmatrix} 2 & 6 & 10 & 0 \\ 0 & 0 & -4 & 4 \\ 0 & 5 & 13 & -8 \end{pmatrix} \Rightarrow \begin{pmatrix} 2 & 6 & 10 & 0 \\ 0 & 5 & 13 & -8 \\ 0 & 0 & -4 & 4 \end{pmatrix}$$

   En utilisant la rementé on trouve:

$$\begin{cases} z=\frac{4}{-4}=-1 \\ y=\frac{1}{5}(-8-13\times(-1))=1 \\ x=\frac{1}{2}(-6\times 1-10\times(-1))=2 \end{cases} \Rightarrow x^* = \begin{pmatrix} 2 \\ 2 \\ -1 \end{pmatrix}$$

3. Méthode de Gauss avec normalisation :Elle consiste à normaliser le pivot:

   On a :

$$(1)\begin{cases} a_{11}x_1+a_{12}x_2+a_{13}x_3=b_1 & (l_1) \\ a_{21}x_1+a_{22}x_2+a_{23}x_3=b_2 & (l_2) \\ a_{31}x_1+a_{32}x_2+a_{33}x_3=b_3 & (l_3) \end{cases}$$

On note par (l_i) la $i^{ème}$ équation du système précedent.
On suppose que $a_{11} \neq 0$,
(l_1) s'écrit : $x_1 + \frac{a_{12}}{a_{11}}x_2 + \frac{a_{13}}{a_{11}}x_3 = \frac{b_1}{a_{11}} \quad (l'_1)$
Si on pose :
$(l'_2) = (l_2) - a_{21}(l'_1)$
et
$(l'_3) = (l_3) - a_{31}(l'_1)$
Alors (1) s'écrit

$$(2)\begin{cases} x_1+\frac{a_{12}}{a_{11}}x_2+\frac{a_{13}}{a_{11}}x_3=\frac{b_1}{a_{11}} & (l'_1) \\ a'_{22}x_2+a'_{23}x_3=b'_2 & (l'_2) \\ a'_{32}x_2+a'_{33}x_3=b'_3 & (l'_3) \end{cases}$$

On suppose que $a'_{22} \neq 0$,
(l'_2) s'écrit $x_2 + a''_{23}x_3 = b''_2 \quad (l''_2)$
Si on pose :
$(l''_3) = (l'_3) - a'_{32}(l''_2)$
si $a''_{33} \neq 0$ on pose $(l'''_3) \quad x_3 = \frac{b''_3}{a''_{33}}$
Alors (2) s'écrit

$$(2) \left\{ \begin{array}{r} x_1 + \frac{a_{12}}{a_{11}}x_2 + \frac{a_{13}}{a_{11}}x_3 = \frac{b_1}{a_{11}} \quad (l'_1) \\ x_2 + a''_{23}x_3 = b''_2 \quad\quad (l''_2) \\ x_3 = \frac{b''_3}{a''_{33}} \quad (l'''_3) \end{array} \right.$$

1. Cette statégie est trés utile pour calculer l'inverse d'une matrice.

Nous pouvons nous demander s'il existe une relation entre la matrice de départ et la matrice triangulaire obtenue. Ce lien existe.

3.4 Factorisation LU

Matriciellement la méthode de Gauss consiste à multiplier la matrice A par la matrice L_1 de telle sorte que l'on ait :

$$\begin{aligned} A_1 &= L_1 A \\ \Rightarrow L_1 &= \begin{pmatrix} 1 & 0 & 0 \\ -\frac{a_{21}}{a_{11}} & 1 & 0 \\ -\frac{a_{31}}{a_{11}} & 0 & 1 \end{pmatrix} \end{aligned}$$

On suppose que $a'_{22} \neq 0$, donc on cherche L_2 de telle sorte que

$$\begin{aligned} A_2 = L_2 A_1 &= \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a'_{22} & a'_{23} \\ 0 & 0 & a''_{33} \end{pmatrix} \\ \Rightarrow L_2 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\frac{a'_{32}}{a'_{22}} & 1 \end{pmatrix} \end{aligned}$$

Ainsi on a : $A_2 = L_2 A_1 = U$ et A_2 est une matrice triangulaire supérieure.

De plus si on pose $A_0 = A$ alors $U = A_2 = L_2 L_1 A_0$ c'est à dire que

$$U = L_2 L_1 A \Leftrightarrow A = L_2^{-1} L_1^{-1} U$$

On a L_1 et L_2 sont des matrices inversibles et triangulaires inférieures donc $L_2 * L_1$ est une matrice inversible et triangulaire inférieure.

De même $L = L_1^{-1} * L_2^{-1}$ est une matrice inversible et triangulaire inférieure

Donc $A = LU$.

Ainsi le systéme linéaire (1) $AX = b$ s'écrit

$$LUX = b \Leftrightarrow \begin{cases} Ly = b \text{ avec } L \text{ matrice triangulaire inférieure} \\ UX = y \text{ avec } U \text{ matrice triangulaire supérieure} \end{cases}$$

En conclusion (à admettre) la méthode de Gauss revient à décomposer la matrice A en un produit de deux (2) matrices triangulaires l'une supérieure U et l'autre inférieure L.

Avec :

$$L = \begin{pmatrix} 1 & & & & \\ l_{21} & 1 & & & \\ l_{31} & l_{32} & 1 & 0 & \\ \vdots & & & & \\ l_{n1} & \cdots & & l_{n,n-1} & 1 \end{pmatrix}$$

où l_{ij} est défini à l'étape (6) de l'algorithme d'élimination et (si l'algorithme d'élimination n'exige pas d'échange de lignes).

Nous ne démontrerons pas cette proposition. Nous nous contenterons de la vérifier sur notre exemple.

Exemple :

$$\begin{pmatrix} 1 & 1 & 0 & 3 \\ 2 & 1 & -1 & 1 \\ 3 & -1 & -1 & 2 \\ -1 & 2 & 3 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 4 & 1 & 0 \\ -1 & -3 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & 0 & 3 & 13 \\ 0 & 0 & 0 & -13 \end{pmatrix}.$$

Il y a une classe importante de matrices pour lesquelles l'élimination peut toujours s'opérer sans échange de lignes (i.e. le pivot a_{jj} n'est jamais nul pendant l'algorithme d'élimination). Ce sont les matrices à diagonale strictement dominante.

Définition: Une matrice A est dite à diagonale strictement dominante si pour tout $i = 1, 2, ..., n$, on a :

$$|a_{ii}| > \sum_{j=1, j\neq i}^{n} |a_{ij}|$$

est vérifiée.

Remarque : Si la matrice est à diagonale strictement dominante alors elle est inversible.

3.4.1 Appplications de la Factorisation LU

Si l'on doit résoudre souvent un système où seul le membre de droite change ou son veut calculer l'inverse d'une matrice, il y a intérêt à effectuer la réduction à la forme triangulaire une fois pour toutes.

En effet, si $A = LU$ on peut résoudre: $Ax = b$ en résolvant $Lz = b$ et $Ux = z$. On a :

$$(1)\ Ax = b \Leftrightarrow \begin{cases} (2)\ Lz = b \\ (3)\ Ux = z \end{cases}$$

Dans ce cas $Ax = LUx = L(Ux) = Lz = b$.

Les systèmes (2) et (3) étant triangulaires, la résolution ne nécessite que l'exécution d'une remontée et d'une descente triangulaire.

Exemple :

$A = LU = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 4 & 1 & 0 \\ -1 & -3 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & 0 & 3 & 13 \\ 0 & 0 & 0 & -13 \end{pmatrix}$, on résoud le système $Ax = \begin{pmatrix} 4 \\ 1 \\ -3 \\ 4 \end{pmatrix}$.

$$Lz = b \Rightarrow \begin{cases} z_1 = 4 \\ 2z_1 + z_2 = 1 \\ 3z_1 + 4z_2 + z_3 = -3 \\ -z_1 - 3z_2 + z_4 = 4 \end{cases} \Rightarrow z = \begin{pmatrix} 4 \\ -7 \\ 13 \\ -13 \end{pmatrix}.$$

$$Ux = z \Rightarrow \begin{cases} -13x_4 = -13 \\ 3x_3 + 13x_4 = 13 \\ -x_2 - x_3 - 5x_4 = -7 \\ x_1 + x_2 + 3x_4 = 4 \end{cases} \Rightarrow x = \begin{pmatrix} -1 \\ 2 \\ 0 \\ 1 \end{pmatrix}.$$

3.5 Mesure des erreurs

L'utilisation d'un calculateur pour implanter les algorithmes étudiés conduira inévitablement à des erreurs. Pour mesurer celles-ci, nous devons mesurer la distance entre le vecteur représentant la solution exacte $x = (x_1, ..., x_n)$ et le vecteur $\hat{x} = (\hat{x}_1, ..., \hat{x}_n)$ représentant la solution approchée. Nous pouvons, pour ce faire, utiliser la "longueur" usuelle de R^n i.e.:

$$\|x\|_2 = \{\sum_1^n x_i^2\}^{\frac{1}{2}}$$

pourtant, dans la pratique on lui préfère souvent la longueur

$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$$

Par exemple si $x = (1, -7, 2, 4)$ alors $\|x\|_\infty = 7$.

Exemple :

Si $x = (1, 1, 1, 1)$ alors $\|x\|_\infty = 1$ si $\hat{x} = (1.01, 1.1, 1, 1)$, on a

$$\|x - \hat{x}\|_\infty = 0.1$$

Considérons alors le système

$$\begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 8 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 32 \\ 23 \\ 33 \\ 31 \end{pmatrix}$$

dont la solution exacte est $x = (1, 1, 1, 1)$.

Si dans le membre de droite nous remplaçons b par:

$$\hat{b} = (32.06; 22.87; 33.07; 30.89)$$

nons obtenons

$$\hat{x} = (9.19; -12.59, 4.49, -1.09)$$

C'est-à-dire qu'une erreur relative de l'ordre de:

$$\frac{\|b - \hat{b}\|_\infty}{\|b\|_\infty} = 3 * 10^{-1}$$

sur b a entraîné une erreur relative de l'ordre de

$$\frac{\|x - \hat{x}\|_\infty}{\|x\|_\infty} = 13.52$$

sur la solution.

Nous devons donc soupçonner que l'application de l'arithmétique finie à la résolution d'un tel système serait désastreuse. L'étude de cette question dépasse le cadre de ce programme.

3.6 Exercices

Série $Ax = b$

Exercice I -

1) On considére le systéme linéaire :

$$(1) \qquad \begin{pmatrix} 1 & 5 \\ 1.0001 & 5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 6 \\ 6.0005 \end{pmatrix}$$

Déterminer la solution X de ce systéme.

2) Dans le systéme précédent, on remplace 6.0005 par 6, déterminer la solution X^* de ce nouveau systéme notée (2).

3) Calculer les erreurs relatives sur les données et sur les résultats.

4) Conclusion.

Exercice II -

Résoudre le systéme linéaire (1):

$$(1) \begin{cases} x + 2y + 3z = 1 \\ 2x + 6y + 10z = 0 \\ 3x + 14y + 28z = -8 \end{cases}$$

1) Par Gauss Classique

2) Par Gauss avec pivotage partiel

3) Par Gauss avec pivotage et mise à l'échelle (i.e. $a_{ii} = 1$).

Exercice III -

1) En arithmétique flottante avec 2 chiffres significatifs ($s = 2$ (s est le nombre de digits)) et arrondi, résoudre par élimination de Gauss, les systémes linéaires (1) et (2).

$$(1) \begin{cases} 0.0001x + y = 3 \\ x + 2y = 5 \end{cases} \qquad (2) \begin{cases} x + 2y = 5 \\ 0.0001x + y = 3 \end{cases}$$

2) Conclusion

Exercice IV -

Soit la matrice $A = \begin{pmatrix} 30 & -20 & -10 \\ -20 & 55 & -10 \\ -10 & -10 & 50 \end{pmatrix}$ et $b = \begin{pmatrix} 1 \\ 5 \\ 2 \end{pmatrix}$

1) Ecrire la matrice A sous la forme LU i.e. trouver L et U (sans pivotage) avec L matrice triangulaire Inférieure et U triangulaire supérieur.

2) En déduire le determinant de A

3) Résoudre par Factorisation LU, le systéme linéaire $AX = b$

Exercice VI -

Soit la matrice $A = \begin{pmatrix} 1 & -1 & 2 \\ -2 & 1 & 1 \\ -1 & 2 & 1 \end{pmatrix}$

1) Ecrire la matrice A sous la forme LU i.e. trouver L et U avec L matrice triangulaire Inférieure et U triangulaire supérieur.

2) Utiliser 1) pour calculer le déterminant de la matrice A.

3) Utiliser 1) pour calculer l'inverse de la matrice A.

[1]

[1] S. El Bernoussi, S. El Hajji et A. Sayah

Chapitre 4

Interpolation polynômiale

4.1 Introduction

Nous abordons dans ce chapitre un nouveau type de problème, faisant intervenir la notion d'approximation d'une fonction.

Cette notion a déjà été rencontrée dans les cours d'analyse.

Exemples :

1) D'aprés la Formule de Taylor à l'ordre 5 de la fonction $\sin(x)$, on a :

$$\forall x \in Vois(0),\ \ \sin(x) \simeq x - \frac{x^3}{3!} + \frac{x^5}{5!} + \sin^{(6)}(\xi)\frac{x^6}{6!} \text{ où } \xi \in Vois(0)$$

On a tronqué la formule de Taylor après l'ordre N (ici 5), on obtient :

au voisinage de 0, une approximation de $\sin(x)$ par un polynôme de degré N (ici 5).

l'erreur commise serait de l'ordre de $\sin^{(6)}(\xi)\frac{x^6}{6!}$ où $\xi \in Vois(0)$

Ainsi avec ce type d'approximation, on a :

- Si $N = 3$, $\sin(0.1) = (0.1) - \frac{(0.1)^3}{3!} = 9.983\,3 \times 10^{-2}$
- Si $N = 5$, $\sin(0.1) = 0.1 - \frac{(0.1)^3}{3!} + \frac{(0.1)^5}{5!} = 9.983\,3 \times 10^{-2}$

Avec le logiciel Maple on a : $\sin(0.1) = 9.983\,3 \times 10^{-2}$

2) Avec les cours d'analyse I et II, on ne connait pas d'expression explicite de $I = \int_0^1 e^{-x^2}dx$

Cependant d'aprés :

- La formule du trapèze $I = \int_0^1 e^{-x^2} dx \simeq \frac{f(0)+f(1)}{2} = \frac{1+e^{-1}}{2} = 0.683\,94$
- La formule de Simpson : $I = \int_0^1 e^{-x^2} dx \simeq \frac{1}{6}\left[f(0) + 4f(\frac{1}{2}) + f(1)\right] = \frac{1}{6}(1 + 4e^{-\frac{1}{4}} + e^{-1}) = 0.747\,18$
- En utilisant la méthode des trapèzes et en subdivisant (partageant) le segment (intervalle) $[0, 1]$ en 10 intervalles egaux, on a : $I = \int_0^1 e^{-x^2} dx \simeq \frac{1}{20}e^{-1} + \frac{1}{10}\sum_{i=1}^{9} e^{-\frac{1}{100}i^2} + \frac{1}{20} = 0.746\,21$
- En utilisant la méthode de Simpson et en subdivisant (partageant) le segment (intervalle) $[0, 1]$ en 10 intervalles égaux, on a : $I = \int_0^1 e^{-x^2} dx \simeq \frac{1}{30}e^{-1} + \frac{1}{15}\sum_{i=1}^{4} e^{-\frac{1}{25}i^2} + \frac{2}{15}\sum_{i=1}^{5} \exp\left(-\left(\frac{1}{5}i - \frac{1}{10}\right)^2\right) + \frac{1}{30} = 0.746\,82$

Avec le Logiciel Maple, on a : $\int_0^1 e^{-x^2} dx = \frac{1}{2}\sqrt{\pi} erf\,(1) = 0.746\,82$

NB : *$erf\,()$ est "The Error Function". Elle est définie pour tout x par : $erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$.*

Donc l'erreur relative (la qualité de l'approximation) dépend du type d'approximation choisie.

On ne connait pas à ce niveau du cours l'expression explicite de l'erreur.

La notion d'approximation d'une fonction consiste à remplacer un problème donné par un problème voisin (un problème majeur en analyse numérique).

La question fondamentale serais de savoir la qualité de cette approximation i.e. la solution (du problème approché) obtenue est -elle aussi voisine que l'on veut de la solution du problème initial.

Remarque : En pratique la fonction f est connue explicitement, ou seulement par ses valeurs en quelques points.

La notion d'interpolation polynomiale est la façon la plus simple d'obtenir une telle approximation.

Théorème : (à admettre)

Soit f une fonction continue dans $[a, b] \subset IR$, alors pour tout $\epsilon > 0$ donné, il existe un polynôme P_n de degré n tel que

$$\underset{x \in [a,b]}{Max} |f(x) - P_n(x)| < \epsilon$$

Ce théorème ne permet pas de construire (de déterminer explicitement) le polynôme P_n. Il existe cependant un certain nombre de techniques (algorithmes) qui le permettent :

1. L'interpolation polynômiale: Elle est la plus classique et est un outil pour la construction des méthodes d'intégration numérique ou des méthodes d'approximation des équations différentielles.

   Remarque : Pour les équations aux dérivées partielles, la méthode des éléments finis, un des outils de base de l'ingénierie moderne, utilise de façon essentielle l'interpolation multi-dimensionnelle.

2. L'interpolation par les fonctions splines : Elle est plus stable que l'interpolation polynômiale, est largement utilisée dans tous les programmes de dessin assisté par ordinateur, conception assistée par ordinateur ou plus généralement de graphisme.

3. Les séries de Fourier et leur analogue discret, la transformation de Fourier discrète : Elles sont un moyen très utile pour l'approximation des fonctions périodiques.

Remarque : L'analyse de Fourier est à la base de nombreuses applications, par exemple en traitement du signal.

Remarque : Une façon naturelle d'approcher les fonctions périodiques est d'utiliser les polynômes trigonométrique.

Nous allons nous limiter à l'introduction de l'interpolation Polynômiale : c'est la façon la plus classique et la plus simple d'approcher une fonction. Elle consiste à déterminer un polynôme $P_n(x)$ de degré n qui puisse remplacer lors des applications la fonction $f(x)$.

De plus, c'est un outil efficace pour :

- Calculer, pour x donné, une approximation de $f(x)$ en calculant $P_n(x)$
- Construire :
  1. des méthodes d'intégration numérique
  2. des méthodes de différentation
  3. des méthodes d'approximation des équations différentielles

4. ...

(nous reviendrons en détails sur ces points dans les chapitres suivants).

Le principe est simple, le procédé est le suivant :

- On choisit (ou on se donne) $(n+1)$ points $x_0, x_1, ..., x_n$.
- On calcule $y_0 = f(x_0), ..., y_n = f(x_n)$
  ou on se donne (x_i, y_i), $i = 0, ..., n$.
- On cherche un polynôme de degré n tel que $P_n(x_i) = y_i$, $i = 0, ..., n$.

Remarque :

1) Les points $(x_i, y_i)_{i=0,n}$ sont appelés points d'interpolation.

2) Si la fonction f est connue seulement par ses valeurs en quelques points, les $(n+1)$ points $x_0, x_1, ..., x_n$ sont fixés..

3) Si on veut que $P_m(x_i) = f(x_i)$ et $P'_m(x_i) = f'(x_i)$, $i = 0, ..., n$, on obtient l'interpolation dite d'Hermite.

La notion d'interpolation polynomiale est la façon la plus simple d'obtenir une telle approximation.

Nous allons montrer l'existence d'un tel polynôme $P_n(x) = a_n x^n + ... + a_0$ en le construisant effectivement.

Il existe plusieurs techniques pour calculer $P_n(x)$. Les plus connues sont celles de Lagrange et de Newton-Côtes. Elles produisent en fin de compte le même résultat. Chaque méthode a ses avantages et ses inconvénients.

Nous allons en fait le faire des deux façons :

1. Une méthode directe basée sur la résolution d'un systéme linéaire
2. Une méthode itérative due à Lagrange.

Nous terminerons ce chapitre par :

1. Une brève discution sur l'erreur d'interpolation polynomiale
2. Une brève description du principe de la méthode itérée de Newton-Côtes

4.2 Une méthode directe basée sur la résolution d'un systéme linéaire:

- On se donne $(n+1)$ points $x_0, x_1, ..., x_n$.
- On calcule $y_0 = f(x_0), ..., y_n = f(x_n)$.
- On cherche un polynôme de degré n tel que $P_n(x_i) = y_i$, $i = 0, ..., n$.

Écrivons explicitement $P_n(x_i) = y_i$.

$$a_n x_i^n + a_{n-1} x_i^{n-1} + ... + a_1 x_i + a_0 = y_i, \qquad i = 0, ..., n$$

On peut réécrire ces $(n+1)$ équations sous forme matricielle :

$$\begin{pmatrix} x_0^n & x_0^{n-1} & \cdots & x_0 & 1 \\ x_1^n & x_1^{n-1} & \cdots & x_1 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_n^n & x_n^{n-1} & \cdots & x_n & 1 \end{pmatrix} \begin{pmatrix} a_n \\ a_{n-1} \\ \vdots \\ a_0 \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{pmatrix}$$

La matrice de ce système est une matrice de type Vandermonde.
On montre que son déterminant est

$$\det = \prod_{i<j} (x_i - x_j)$$

On a $\det \neq 0$ si tous les x_i sont distincts. On peut donc trouver un unique vecteur de coefficients $(a_n, ..., a_0)$ résolvant le problème.

Il est connu (à admettre) que les matrices du type Vandermonde deviennent très mal conditionnées lorsque n augmente (elle sont trés sensible aux erreurs d'arrondies).

Dans la pratique, cette méthode n'est à utiliser que si $n \leq 3$. Il serait à la fois inutile et dangereux de vouloir l'utiliser pour n grand.

4.3 Une méthode itérative : Méthode de Lagrange

4.3.1 Interpolation Linéaire :

On considère deux points $(x_0, y_0), (x_1, y_1)$ avec :

$$\begin{cases} x_0 \neq x_1 \\ y_0 = f(x_0) \text{ et } y_1 = f(x_1). \end{cases}$$

Pour déterminer le polynôme $P_1(x)$ de dégré 1 (d'équation : $y = ax + b$) qui passe par deux points distincts $(x_0, y_0), (x_1, y_1)$ $(x_0 \neq x_1)$. On peut:

1) Résoudre le système d'équations:

$$\begin{cases} ax_0 + b = y_0 \\ ax_1 + b = y_1 \end{cases}$$

d'où

$$\begin{cases} a = \frac{(y_1 - y_0)}{(x_1 - x_0)} \\ b = y_0 - ax_0 = \frac{x_1 y_0 - x_0 y_1}{x_1 - x_0} \end{cases}$$

On a :

$$P_1(x) = \frac{(y_1 - y_0)}{(x_1 - x_0)} x + (\frac{x_1 y_0 - x_0 y_1}{x_1 - x_0})$$

et

$$P_1(x_0) = y_0 \text{ et } P_1(x_1) = y_1$$

2) Poser

$$L_0(x) = \frac{x - x_1}{x_0 - x_1}$$
$$L_1(x) = \frac{x - x_0}{x_1 - x_0}$$

On a:

$$L_k(x_i) = \begin{cases} 0 & \text{si} \quad i \neq k \\ 1 & \text{si} \quad i = k \end{cases}$$

Ainsi,

$$\begin{aligned}P_1(x) &= y_0L_0(x) + y_1L_1(x)\\ &= y_0\frac{(x-x_1)}{(x_0-x_1)} + y_1(\frac{x-x_0}{x_1-x_0})\\ &= \frac{(y_1-y_0)}{(x_1-x_0)}x + (\frac{x_1y_0-x_0y_1}{x_1-x_0})\end{aligned}$$

On a :

$$P_1(x_0) = y_0 \text{ et } P_1(x_1) = y_1$$

car

$$L_k(x_i) = \begin{cases} 0 & \text{si} \quad i \neq k \\ 1 & \text{si} \quad i = k \end{cases}$$

Ces deux procédés déterminent évidemment le même polynôme de dégré 1 (la même droite).

Si maintenant, on veut déterminer le polynome de degré 2 qui passe par trois (3) points distincts alors:

i) la première expression de $P_1(x)$ est inadéquate (il faut refaire les calculs)

ii) la dexième expression se prête assez facilement à une généralisation par récurrence.

Exemple :

Déterminer le polynôme d'interpolation $P_1(x)$ de degré 1 tel que
$P_1(x_i) = f(x_i)$, $i = 0, 1$
avec $y_i = f(x_i)$ $i = 0, 1$, $(x_0, y_0) = (0, 1), (x_1, y_1) = (2, 5)$

On a déterminé le polynôme d'interpolation qui passe par les 2 points : $(0, 1)$ et $(2, 5)$

D'aprés la méthode de Lagrange,

$$\begin{aligned}P_1(x) &= y_0L_0(x) + y_1L_1(x)\\ &= y_0\frac{(x-x_1)}{(x_0-x_1)} + y_1(\frac{x-x_0}{x_1-x_0})\\ &= 1\frac{(x-2)}{(0-2)} + 5\frac{(x-0)}{(2-0)}\\ &= 2x+1\end{aligned}$$

4.3.2 Interpolation parabolique

On considére trois points $(x_0, y_0), (x_1, y_1)$ et (x_2, y_2) avec :

$$\left\{ \begin{array}{c} x_0 \neq x_1 \text{ ,et } x_0 \neq x_2 \quad et \quad x_1 \neq x_2 \\ y_0 = f(x_0),\ y_1 = f(x_1) \text{ et } y_2 = f(x_2). \end{array} \right.$$

Pour déterminer le polynôme $P_2(x)$ de dégré 2, d'équation $y = ax^2+bx+c$ qui passe par trois points distincts $(x_0, y_0), (x_1, y_1)$ et (x_2, y_2), il suffit de poser:

$$\begin{aligned} L_0(x) &= \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} \\ L_1(x) &= \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} \\ L_2(x) &= \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)} \end{aligned}$$

On a :

$$L_k(x_i) = \left\{ \begin{array}{lll} 0 & \text{si} & i \neq k \\ 1 & \text{si} & i = k \end{array} \right.$$

Ainsi

$$\begin{aligned} P_2(x) &= y_0 L_0(x) + y_1 L_1(x) + y_2 L_2(x) \\ &= y_0 \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} + y_1 \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} + y_2 \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)} \end{aligned}$$

est le polynôme d'interpolation polynômiale associé.

Exemple :

Déterminer le polynôme d'interpolation $P_2(x)$ de degré 2 tel que $P_2(x_i) = f(x_i)$, $i = 0, 1$ et 2

avec $y_i = f(x_i)$ $i = 0, 1$ et 2 , $(x_0, y_0) = (0, 1), (x_1, y_1) = (1, 2)$ et $(x_2, y_2) = (2, 5)$

On a déterminé le polynôme d'interpolation qui passe par les 3 points : $(0, 1), (1, 2)$ et $(2, 5)$

D'aprés la méthode de Lagrange,

$$\begin{aligned}
P_2(x) &= y_0L_0(x) + y_1L_1(x) + y_2L_2(x) \\
&= y_0\frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} + y_1\frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} + y_2\frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)} \\
&= 1\frac{(x-1)(x-2)}{(-1)(-2)} + 2\frac{(x)(x-2)}{(1)(-1)} + 5\frac{(x)(x-1)}{(2)(1)} \\
&= x^2+1
\end{aligned}$$

Remarque :

1) Pour calculer $P_2(x)$,on n'a pas utilisé le polynome $P_1(x)$ calculé dans l'exemple précédent et pourtant on avait deux points communs.

2) $L_i(x), i = 0, 1, 2$ sont des polynômes de degré 2 :

$L_0(x) = \frac{(x-1)(x-2)}{(-1)(-2)} = \frac{1}{2}(x-1)(x-2) = \frac{1}{2}x^2 - \frac{3}{2}x + 1$

$L_1(x) = \frac{(x)(x-2)}{(1)(-1)} = -x(x-2) = -x^2 + 2x$

$L_2(x) = \frac{(x)(x-1)}{(2)(1)} = \frac{1}{2}x(x-1) = \frac{1}{2}x^2 - \frac{1}{2}x$

On considére $(L_i(x))_{i=0,2}$ comme une base de l'interpolation polynômiale quadratique

Dans l'intervallle $[0, 2]$, il existe plusieurs fonctions $f(x)$ qui passent par les 3 points $(x_0, y_0) = (0, 1), (x_1, y_1) = (1, 2)$ et $(x_2, y_2) = (2, 5)$ mais elle ne seront pas approchées par $P_2(x) = x^2 + 1$ de la même façon.

4.3.3 Interpolation de Lagrange

- On choisit $n+1$ points $x_0, x_1, ..., x_n$.
- On calcule $y_0 = f(x_0), ..., y_n = f(x_n)$.
- On cherche un polynôme de degré n tel que $P_n(x_i) = y_i,\ i = 0, ..., n$.

On introduit les coefficients d'interpolation de Lagrange.

$$L_k(x) = \frac{(x-x_0)...(x-x_{k-1})(x-x_{k+1})...(x-x_n)}{(x_k-x_0)...(x_k-x_{k-1})(x_k-x_{k+1}...(x_k-x_n)}$$

$$L_k(x) = \prod_{j=0\ j\neq k}^{j=n} \frac{(x-x_j)}{(x_k-x_j)}$$

$L_k(x)$ est un polynôme de degré n,

$$L_k(x_i) = \begin{cases} 0 & \text{si} \quad i \neq k \\ 1 & \text{si} \quad i = k \end{cases}$$

Donc

$$P(x) = y_0 L_0(x) + y_1 L_1(x) + ... + y_n L_n(x) = \sum_{k=0}^{n} y_k L_k(x)$$

est un polynôme de degré n qui vérifie bien $P(x_i) = y_i$

Propriété : Le Polynôme d'interpolation polynômiale est unique.

En effet si $P(x)$ et $Q(x)$ sont deux polynômes d'interpolation alors :

$P(x) - Q(x)$ est un polynôme de degré n pour lequel

$$P(x_i) - Q(x_i) = 0, \qquad i = 0, ..., n.$$

Ce polynôme de degré $\leq n$ ayant $n+1$ racines, il est identiquement nul.

Exemple :

On suppose que $f(x) = \sqrt[3]{x}$ et que $(x_0, y_0) = (0,0), (x_1, y_1) = (1,1)$ et $(x_2, y_2) = (8,2)$

1) Déterminer le polynôme $P_2(x)$ d'interpolation polynômiale qui passent par les points $(x_i, y_i)_{i=0,2}$

On a à déterminer le polynôme d'interpolation qui passe par les 3 points : $(0,0), (1,1)$ et $(8,2)$

D'aprés la méthode de Lagrange,

$$\begin{aligned} P_2(x) &= y_0 L_0(x) + y_1 L_1(x) + y_2 L_2(x) \\ &= y_0 \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} + y_1 \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} + y_2 \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)} \\ &= 0 \frac{(x-1)(x-2)}{(0-1)(0-2)} + 1 \frac{(x-0)(x-8)}{(1-0)(1-8)} + 2 \frac{(x-0)(x-1)}{(8-0)(8-1)} \end{aligned}$$

$$P_2(x) = -\frac{3}{28}x^2 + \frac{31}{28}x$$

On a bien $P_2(0) = 0, P_2(1) = 1$ et $P_2(8) = -\frac{3}{28}(8)^2 + \frac{31}{28}8 = 2$

2) Calculer $P_2(x)$ et $f(x) = \sqrt[3]{x}$ pour $x = 0.5, 0.95, 1, 1.5$ et 3. Conclusion.

On a :

x	$f(x)$	$P_2(x) = -\frac{3}{28}x^2 + \frac{31}{28}x$
0.5	0.7937	0.52679
0.95	0.98305	0.95509
1	1	1
1.5	1.1447	1.4196
3	1.4422	$\frac{33}{14} = 2.3571$

L'interpolation polynomiale de degré 2 ne fournit de résultat acceptable qu'au voisinage des points d'interpolation ici 1.

3) Tracer le graphe de $f(x)$ et $P_2(x)$.Conclusion.

On voit que dans l'intervalle $[2, 6]$, $P_2(x)$ fournit une mauvaise approximation de $f(x)$.

Pour x donne, $P_2(x)$ fournira une bonne approximation de $f(x)$ si x est voisin de $0, 1$ et 8.

Remarque :

1) En pratique, on utilise l'interpolation polynômiale avec des polynômes de dégré n assez grand ou l'interpolation polynômiale par morceaux. Ainsi dans l'exemple précedent, il faut augmenter le nombre de points d'interpolations.

2) Si les valeurs y_k sont des valeurs expérimentales. L'interpolation polynomiale est une technique peu appropriée pour de telles situations. Les polynômes de degré élevé sont sensibles à la perturbation des données.

3) La méthode de Lagrange s'adapte mal au changement du nombre de points $(x_i, y_i)_i$. On ne peut utiliser les coefficients de Lagrange si on passe de n à $(n + 1)$ points.

4) **Phénomène de RUNGE** (fonction de Runge) : L'interpolation polynômiale ne fournit pas une bonne approximation de la fonction $f(x) = \frac{1}{1+25x^2}$. Si on augmente le nombre de points d'interpolation le resultat devient plus mauvais. (A admettre).

4.4 Interpolation Itérée de Newton-Côtes

- On choisit $n + 1$ points $x_0, x_1, ..., x_n$.
- On calcule $y_0 = f(x_0), ..., y_n = f(x_n)$.
- On cherche un polynôme de degré n tel que $P_n(x_i) = y_i$, $i = 0, ..., n$.

L'Interpolation Itérée de Newton-Côtes est un procédé itératif qui permet de calculer le polynôme d'interpolation $P_n(x)$ de dégré n basé sur $(n+1)$ points $(x_i, y_i)_{i=0,n}$ à partir du polynôme d'interpolation $P_{(n-1)}(x)$ de dégré $(n-1)$ basé sur n points $(x_i, y_i)_{i=0,(n-1)}$, en posant :

$$P_n(x) = P_{(n-1)}(x) + C(x), \; n \geq 1$$

avec

$$C(x) = a_n(x - x_0)(x - x_1)...(x - x_{(n-1)})$$
$$a_n = \sum_{k=0}^{n} \frac{f(x_k)}{(x_k - x_0)...(x_k - x_{(k-1)})(x_k - x_{(k+1)})...(x_k - x_n)}$$

Les coéfficients a_n sont appelés différences divisées d'ordre n de la fonction f, on note :

$$a_n = f\left[x_0, x_1, ..., x_n\right]$$

- On appelle "différence divisée d'ordre 0 de f en un point x" la valeur définie par
$$f\left[x\right] = f(x)$$
- Différence "divisée d'ordre 1 de f en deux points x et y" la valeur définie par
$$f\left[x, y\right] = \frac{f\left[x\right] - f\left[y\right]}{x - y}$$
on a
$$f\left[x, y\right] = \frac{f(x)}{x - y} + \frac{f(y)}{y - x}$$
- Différence "divisée d'ordre 2 de f en deux points x, y et z" la valeur définie par
$$f\left[x, y, z\right] = \frac{f\left[x, y\right] - f\left[y, z\right]}{x - z}$$
$$= \frac{f(x)}{(x - y)(x - z)} + \frac{f(y)}{(y - x)(y - z)} + \frac{f(z)}{(z - x)(z - y)}$$

et plus généralement:

$$f[x_1, x_2, ..., x_n] = \sum_{i=1}^{n} \frac{f(x_i)}{\prod\limits_{\substack{k=1 \\ k \neq i}}^{n} (x_i - x_k)}$$

Remarque:
Les différences divisées sont indépendants de l'ordre des points.

Quel est le lien entre $f(x)$ et lex différences divisées?
Soit x un point autre que les $n+1$ points x_i, $i = 1, ..., n$. On a

$$f[x, x_0] = \frac{f(x) - f[x_0]}{x - x_0}$$

$d'où$

$$f(x) = f[x_0] + (x - x_0) f[x, x_0]$$

mais comme

$$f[x, x_0, x_1] = \frac{f[x, x_0] - f[x_0, x_1]}{x - x_1}$$

alors

$$f(x) = f[x_0] + (x - x_0) f[x_0, x_1] - (x - x_0)(x - x_1) f[x, x_0, x_1]$$

en continuant ainsi de proche en proche on obtient:

$$f(x) = f[x_0] + (x - x_0) f[x_0, x_1] + ... + (x - x_0) ... (x - x_{n-1}) f[x_0, ..., x_n] + (x - x_0)...(x - x_n) f[x, x_0, ..., x_n]$$

on vérifie que

$$f(x) = P_n(x) + L(x) f[x, x_0, ..., x_n]$$

où $P_n(x)$ est un polynôme de degré n tel que $P_n(x_i) = f(x_i)$, pour $i = 0, ..., n$. C'est donc le polynôme d'interpolation de Lagrange, on l'appelle le polynôme de Newton.

Comme signalé dans l'introduction, l'interpolation polynomiale sera utilisé comme outil d'approximation (pour la construction des méthodes d'intégration numérique ou des méthodes de dérivation numérique ou des méthodes d'approximation des équations différentielles), il est donc fondamental de connaitre une expression de l'erreur d'interpolation.

4.5 Erreur d'Interpolation polynomiale :

L'erreur commise lors d'une interpolation est une question fondamentale en analyse numérique:

- elle renseigne à priori sur la nature de cette erreur
- elle fournit des informations sur les termes qui y participent
- elle permet d'avoir un ordre de grandeur de l'erreur commise.

Nous allons énoncer un résultat qui répond à ces interrogations dans le cas où la fonction f est régulière (de classe C^p, p assez grand).

Théorème :

Soient f une fonction de classe C^{n+1} dans I et , $(x_i)_{i=0,n}$ $(n+1)$ points distincts dans I avec $x_0 < x_1 < ... < x_n$

Alors pour tout $x \in [x_0, x_n]$, il existe $\zeta = \zeta(x)$ tel que

$$f(x) - P_n(x) = \frac{f^{(n+1)}(\zeta)}{(n+1)!}(x - x_0)(x - x_1)...(x - x_n) = \frac{f^{(n+1)}(\zeta)}{(n+1)!}L(x)$$

où

$$P_n(x) = y_0L_0(x) + y_1L_1(x) + ... + y_nL_n(x) = \sum_{k=0}^{n} y_kL_k(x)$$

avec $$L_k(x) = \prod_{j=0\ j\neq k}^{n} \frac{(x - x_j)}{(x_k - x_j)}$$

et $$L(x) = (x - x_0)(x - x_1)...(x - x_n)$$

$P_n(x)$ est le polynôme d'interpolation de Lagrange.

Remarque :

1) Cette formule montre que :

i) l'erreur est nulle pour $x = x_i$ i.e. x est un point d'interpolation.

ii) l'erreur dépend de la fonction considérée (de $f^{(n+1)}$) et des points d'interpolations $(x_i)_i$.

2) Cette formule d'erreur permet de trouver des formules d'erreur pour l'intégration numérique et la differentiabilité numérique.

Dans le cas de l'erreur d'interpolation à partir de la forme de Newton, on a:

$$f(x) - P_n(x) = L(x).f[x, x_0, ..., x_n].$$

Comme on a la même fonction f selon les mêmes points x_i pour $i = 0, ..., n$, il s'agit de deux formes du même polynôme, et l'erreur d'interpolation est la même, d'où

$$f(x) - P_n(x) = \frac{f^{(n+1)}(\zeta)}{(n+1)!}L(x) = L(x).f[x, x_0, ..., x_n].$$

Exemple :

Déterminer l'erreur d'interpolation polynomiale dans le cas de l'interpolation parabolique

On approche la fonction $f(x)$ par la parabole passant par les points $(x_0, y_0) = (0, 1), (x_1, y_1) = (1, 2)$ et $(x_2, y_2) = (2, 5)$.

Le polynôme d'interpolation $P_2(x)$ de degré 2 tel que $P_2(x_i) = f(x_i)$, $i = 0, 1$ et 2

avec $y_i = f(x_i)$ $i = 0, 1$ et 2 , $(x_0, y_0) = (0, 1), (x_1, y_1) = (1, 2)$ et $(x_2, y_2) = (2, 5)$

D'aprés la méthode de Lagrange,

$$\begin{aligned} P_2(x) &= y_0L_0(x) + y_1L_1(x) + y_2L_2(x) \\ &= 1\frac{(x-1)(x-2)}{(-1)(-2)} + 2\frac{(x)(x-2)}{(1)(-1)} + 5\frac{(x)(x-1)}{(2)(1)} \\ &= x^2 + 1 \end{aligned}$$

D'après le théorème précédent,

$$\begin{aligned} f(x) - P_2(x) &= \frac{f^{(3)}(\zeta)}{3!}(x - x_0)(x - x_1)(x - x_2) \\ &= \frac{f^{(3)}(\zeta)}{3!}x(x-1)(x-2) \end{aligned}$$

Si $\left|f'^{(3)}(x)\right| \leq M$ alors

$$
\begin{aligned}
\forall x \in [0,2] \ , \quad & |f(x) - P_2(x)| \\
|f(x) - P_2(x)| &\leq \frac{M}{6}\,|x(x-1)(x-2)| \\
&\leq \frac{M}{6} x(x-1)(x-2) \\
&\leq 6.4 * 10^{-2} * M.
\end{aligned}
$$

(le maximum de $u(x) = x(x-1)(x-2)$ est atteint en $x^* = \frac{3-\sqrt{3}}{3}$; d'où $\frac{1}{6}u(x^*) = \frac{1}{6}\frac{3-\sqrt{3}}{3}\left(\frac{3-\sqrt{3}}{3} - 1\right)\left(\frac{3-\sqrt{3}}{3} - 2\right) = 0.0\,641\,5 \sim 6.4 * 10^{-2}$).

4.6 Exercices:

Série Interpolation Numérique

Exercice I :

1) Déterminer par une méthode directe basée sur la résolution d'un systéme linéaire, le polynôme d'interpolation $P_1(x)$ de degré 1 tel que $P_1(x_i) = f(x_i)$, $i = 0, 1$ avec $y_i = f(x_i)$ $i = 0, 1$, $(x_0, y_0) = (-2, 4)$ et $(x_1, y_1) = (2, 8)$

2) Déterminer par une méthode directe basée sur la résolution d'un systéme linéaire, le polynôme d'interpolation $P_2(x)$ de degré 2 tel que $P_2(x_i) = f(x_i)$, $i = 0, 1$ et 2 avec $y_i = f(x_i)$ $i = 0, 1$ et 2, $(x_0, y_0) = (-2, 4), (x_1, y_1) = (0, 2)$ et $(x_2, y_2) = (2, 8)$. Conclusion.

Exercice II :

1) Déterminer par la méthode de Lagrange, le polynôme d'interpolation $P_1(x)$ de degré 1 tel que $P_1(x_i) = f(x_i)$, $i = 0, 1$ où $y_i = f(x_i)$ $i = 0, 1$, $(x_0, y_0) = (-2, 4)$ et $(x_1, y_1) = (2, 8)$

2) Déterminer par la méthode de Lagrange, le polynôme d'interpolation $P_2(x)$ de degré 2 tel que $P_2(x_i) = f(x_i)$, $i = 0, 1$ et 2 où $y_i = f(x_i)$ $i = 0, 1$ et 2, $(x_0, y_0) = (-2, 4), (x_1, y_1) = (0, 2)$ et $(x_2, y_2) = (2, 8)$

Exercice III :

1) Déterminer par la méthode de Newton-Côtes, le polynôme d'interpolation $P_1(x)$ de degré 1 tel que $P_1(x_i) = f(x_i)$, $i = 0, 1$ où $y_i = f(x_i)$ $i = 0, 1$, $(x_0, y_0) = (-2, 4)$ et $(x_1, y_1) = (2, 8)$.

2) Déterminer par la méthode de de Newton, le polynôme d'interpolation $P_2(x)$ de degré 2 tel que $P_2(x_i) = f(x_i)$, $i = 0, 1$ et 2 où $y_i = f(x_i)$ $i = 0, 1$ et 2, $(x_0, y_0) = (-2, 4), (x_1, y_1) = (0, 2)$ et $(x_2, y_2) = (2, 8)$. Conclusion.

Exercice IV :

On suppose que $(x_0, y_0) = (0, 0), (x_1, y_1) = (1, 1)$ et $(x_2, y_2) = (2, 8)$

1) Déterminer par la méthode de Lagrange, le polynôme d'interpolation $P_2(x)$ de degré 2 tel que $P_2(x_i) = y_i; i = 0, 1, 2$.

2) Tracer le graphe des fonctions $P_2(x) = 3x^2 - 2x$ et $f(x) = x^3$ dans l'intervalle $[0, 2]$.

3) Calculer $P_2(x)$ et $f(x) = x^3$ pour $x = 0.9, 1.1, 1.99, 2.1$ et 5. Conclusion.

4) Déterminer l'erreur commise si on en remplace dans l'intervalle $[0, 2]$, $f(x) = x^3$ par $P_2(x) = 3x^2 - 2x$.

Exercice V :

On suppose que $(x_0, y_0) = (0, 1), (x_1, y_1) = (0.5, e^{\frac{1}{2}})$, $(x_2, y_2) = (1, e)$

1) Déterminer par la méthode de Lagrange, le polynôme d'interpolation $P_2(x)$ de degré 2 tel que $P_2(x_i) = y_i$, $i = 0, 1, 2$ et 3.

2) i) Déterminer une expression de l'erreur d'interpolation polynomiale.

ii) Déterminer une borne de l'erreur d'interpolation polynomiale. Indépendantes de ξ où $\xi = \xi(x)$.

ii) Déterminer une borne de l'erreur d'interpolation polynomiale. Indépendantes de ξ et de x.

[1]

[1] S. El Bernoussi, S. El Hajji et A. Sayah

Chapitre 5

Integration et dérivation numérique.

5.1 Introduction :

Si f est une fonction dérivable sur $[a, b]$, la dérivée en $c \in \,]a, b[$ est définie par:

$$f'(c) = \lim_{h \to 0} \frac{\Delta f(c)}{h}$$
$$\text{où } \Delta f(c) = f(c+h) - f(c)$$

Si f est une fonction continue sur $[a, b]$, l'integrale de f sur $[a, b]$ est définie par

$$\int_a^b f(x)dx = \lim_{h \to 0} R(h)$$
$$\text{où } R(h) = \sum_{k=1}^{n} f(a+kh).h$$

$R(h)$ est la somme de Riemann avec $h = \frac{b-a}{n}$.

On sait déterminer $f'(c)$ "exactement" pour f définie à partir de fonctions élémentaires (exp: $sinx$, $e^x, \ln x, ...$).

On sait aussi calculer $\int_a^b f(x)dx$ en utilisant les théorèmes fondamentaux d'intégration pour une fonction continue sur $[a, b]$, et on a $\int_a^b f(x)dx = F(b) - F(a)$ où $F(x)$ est une primitive de $f(x)$.

Mais il existe des fonctions très simples comme $\frac{\sin x}{x}$ ou $\sqrt{\cos^2 x + 3\sin^2 x}$ qui n'ont pas de primitive connue, donc, comment peut-on integrer de telles fonctions entre a et b?

D'autre part f peut-être connue seulement en quelques points et sa formule est inconnue (exp: résultats experimentaux,...), donc comment peut-on dériver ou intégrer ses fonctions?

Du point de vue numérique, la solution à ce problème est immédiate: nous avons vu, dans les chapitres précédents, comment approximer une fonction par une fonction plus simple, facile à dériver ou à intégrer.

De façon précise si $P(x)$ est une approximation de f dans l'intervalle $[a, b]$, nous nous proposons d'étudier les approximations:

$$f'(y) \approx P'(y) \qquad y \in [a, b]$$

$$et$$

$$\int_a^b f(x)dx \approx \int_a^b P(x)dx.$$

5.2 Dérivation.

La dérivation numérique nous permet de trouver une estimation de la dérivée ou de la pente d'une fonction, en utilisant seulement un ensemble discret de points.

5.2.1 Dérivée première.

Soit f une fonction connue seulement par sa valeur en $(n+1)$ points donnés $x_i \;\; i = 0, 1, ..., n$ distincts.

Les formules de différence les plus simples basées sur l'utilisation de la ligne droite pour interpoler les données ulilisent deux points pour estimer la dérivée.

On suppose connue la valeur de la fonction en x_{i-1}, x_i et x_{i+1}; on pose $f(x_{i-1}) = y_{i-1}, f(x_i) = y_i$ et $f(x_{i+1}) = y_{i+1}$.

Si on suppose que l'espace entre deux points successifs est constant, donc on pose $h = x_i - x_{i-1} = x_{i+1} - x_i$.

Alors les formules standarts en deux points sont:

Formule de difference progressive :

$$f'(x_i) \approx \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} = \frac{y_{i+1} - y_i}{x_{i+1} - x_i}.$$

Formule de difference régressive :

$$f'(x_i) \approx \frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}} = \frac{y_i - y_{i-1}}{x_i - x_{i-1}}.$$

Formule de difference centrale

$$f'(x_i) \approx \frac{f(x_{i+1}) - f(x_{i-1})}{x_{i+1} - x_{i-1}} = \frac{y_{i+1} - y_{i-1}}{x_{i+1} - x_{i-1}}.$$

Les trois formules classiques de différences sont visualisées sur la figure suivante, et sont les conséquences de la définition de la dérivée:

Exemple :

Pour illustrer les trois formules, considérons les données suivantes:

$(x_0, y_0) = (1, 2); (x_1, y_1) = (2, 4); (x_2, y_2) = (3, 8); (x_3, y_3) = (4, 16)$ et $(x_4, y_4) = (5, 32)$.

Nous voulons estimer la valeur de $f'(x_2)$.

Progressive: $f'(x) \approx \frac{f(x_3)-f(x_2)}{x_3-x_2} = \frac{16-8}{4-3} = 8.$

Regressive : $f'(x) \approx \frac{f(x_2)-f(x_1)}{x_2-x1} = \frac{8-4}{3-2} = 4.$

Centrale : $f'(x) \approx \frac{f(x_3)-f(x_1)}{x_3-x1} = \frac{16-4}{4-2} = 6.$

Les données ont été calculé pour la fonction $f(x) = 2^x$. $f'(x) = 2^x \ln(2)$ et pour $x = 3$ $f'(3) = 2^3 \ln(2) = 5.544$.

Remarque:

Les formules de différences classiques peuvent être trouvées en utilisant la formule de Taylor.

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2} f''(\eta).$$
$$x \leq \eta \leq x + h$$

- Formule progressive:

$$h = x_{i+1} - x_i$$
$$f'(x_i) = \frac{f(x_{i+1}) - f(x_i)}{h} - \frac{h}{2} f''(\eta)$$
$$x_i \leq \eta \leq x_{i+1}$$

l'erreur est $\frac{h}{2}f''(\eta)$ donc en $O(h)$. Cette formule peut être trouvée aussi en utilisant le polynôme d'interpolation de Lagrange pour les points $(x_i, f(x_i))$ et $(x_{i+1}, f(x_{i+1}))$.

- Formule regressive:

$$\begin{aligned} h &= x_i - x_{i-1} \\ f'(x_i) &= \frac{f(x_i) - f(x_{i-1})}{h} + \frac{h}{2}f''(\eta) \\ x_{i-1} &\leq \eta \leq x_i \end{aligned}$$

La formule de différence centrale de la dérivée en x_i peut être trouvée en utilisant la formule de Taylor d'ordre 3 avec $h = x_{i+1} - x_i = x_i - x_{i-1}$

$$\begin{aligned} f(x_{i+1}) &= f(x_i) + hf'(x_i) + \frac{h^2}{2}f''(x_i) + \frac{h^3}{3!}f'''(\eta_1) \\ f(x_{i-1}) &= f(x_i) - hf'(x_i) + \frac{h^2}{2}f''(x_i) - \frac{h^3}{3!}f'''(\eta_2) \\ x_i &\leq \eta_1 \leq x_{i+1}, \ x_{i-1} \leq \eta_2 \leq x_i \end{aligned}$$

si on suppose que f''' est continue sur $[x_{i-1}, x_{i+1}]$ on peut ecrire la formule suivante:

$$\begin{aligned} f'(x_i) &= \frac{f(x_{i+1}) - f(x_{i-1})}{2h} + \frac{h^2}{6}f'''(\eta) \\ x_{i-1} &\leq \eta \leq x_{i+1} \end{aligned}$$

l'erreur est $\frac{h^2}{6}f'''(\eta)$ donc en $O(h^2)$. La formule de différence centrale peut aussi être trouvée à partir du polynôme d'intérpolation de Lagrange en 3 points.

On peut interpoler les données par un polynome au lieu d'utiliser la droite, nous obtenons alors les formules de différence qui utilisent plus de deux points. On suppose que le pas h est constant.

Formule de différence progressive utilisant trois points:

$$f'(x_i) \approx \frac{-f(x_{i+2}) + 4f(x_{i+1}) - 3f(x_i)}{x_{i+2} - x_i}$$

Formule de différence régressive utilisant trois points:

$$f'(x_i) \approx \frac{3f(x_i) - 4f(x_{i-1}) + f(x_{i-2})}{x_i - x_{i-2}}$$

Exemple : Formules de différence en trois points:

En utilisant les données de l'exemple précédent, on trouve:

$f'(x_i) \approx \frac{-32+4(16)-3(8)}{2} = 4$ progressive.

$f'(x_i) \approx \frac{3(8)-4(4)+2}{2} = 5$ regressive.

5.2.2 Formule générale en trois points.

La formule d'approximation en 3 points de la dérivée première, basée sur le polynôme d'interpolation de Lagrange, n'utilise pas des points équidistants.

Etant donné trois points $(x_1, y_1); (x_2, y_2)$ et (x_3, y_3) avec $x_1 < x_2 < x_3$, la formule suivante permet d'approcher la dérivée en un point $x \in [x_1, x_3]$. Les dérivées aux points x_i sont les suivantes:

$$f'(x_1) = \frac{2x_1 - x_2 - x_3}{(x_1 - x_2)(x_1 - x_3)}y_1 + \frac{x_1 - x_3}{(x_2 - x_1)(x_2 - x_3)}y_2 + \frac{x_1 - x_2}{(x_3 - x_1)(x_3 - x_2)}y_3;$$

$$f'(x_2) = \frac{x_2 - x_3}{(x_1 - x_2)(x_1 - x_3)}y_1 + \frac{2x_2 - x_1 - x_3}{(x_2 - x_1)(x_2 - x_3)}y_2 + \frac{x_2 - x_1}{(x_3 - x_1)(x_3 - x_2)}y_3;$$

$$f'(x_3) = \frac{x_3 - x_2}{(x_1 - x_2)(x_1 - x_3)}y_1 + \frac{x_3 - x_1}{(x_2 - x_1)(x_2 - x_3)}y_2 + \frac{2x_3 - x_2 - x_1}{(x_3 - x_1)(x_3 - x_2)}y_3;$$

Le polynôme de Lagrange est donnée par

$$P(x) = L_1(x)y_1 + L_2(x)y_2 + L_3(x)y_3$$

où

$$L_1(x) = \frac{(x - x_2)(x - x_3)}{(x_1 - x_2)(x_1 - x_3)}$$

$$L_2(x) = \frac{(x - x_1)(x - x_3)}{(x_2 - x_1)(x_2 - x_3)}$$

$$L_3(x) = \frac{(x - x_1)(x - x_2)}{(x_3 - x_1)(x_3 - x_2)}$$

L'approximation de la dérivée première est donnée par $f'(x) \approx P'(x)$, qui

peut s'ecrire

$$P'(x) = L_1'(x)y_1 + L_2'(x)y_2 + L_3'(x)y_3$$

où

$$L_1'(x) = \frac{2x - x_2 - x_3}{(x_1 - x_2)(x_1 - x_3)}$$
$$L_2'(x) = \frac{2x - x_1 - x_3}{(x_2 - x_1)(x_2 - x_3)}$$
$$L_3'(x) = \frac{2x - x_1 - x_2}{(x_3 - x_1)(x_3 - x_2)}$$

donc

$$f'(x) = \frac{2x - x_2 - x_3}{(x_1 - x_2)(x_1 - x_3)}y_1 + \frac{2x - x_1 - x_3}{(x_2 - x_1)(x_2 - x_3)}y_2 + \frac{2x - x_1 - x_2}{(x_3 - x_1)(x_3 - x_2)}y_3.$$

5.2.3 Dérivées d'ordre supérieur.

Les formules de dérivées d'ordre supérieur, peuvent être trouvées à partir des dérivées du polynôme de Lagrange ou en utilisant les formules de Taylor.

Par exemple, étant donné 3 points x_{i-1}, x_i, x_{i+1} équidistants, la formule de la dérivée seconde est donnée par:

$$f''(x_i) = \frac{1}{h^2}[f(x_{i+1}) - 2f(x_i) + f(x_{i-1})]$$

l'erreur est en $O(h^2)$.

Dérivée seconde à partir du polynôme de Taylor.

$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \frac{h^3}{3!}f'''(x) + \frac{h^4}{4!}f^{(4)}(\eta_1)$

$f(x-h) = f(x) - hf'(x) + \frac{h^2}{2}f''(x) - \frac{h^3}{3!}f'''(x) + \frac{h^4}{4!}f^{(4)}(\eta_2)$

$x \leq \eta_1 \leq x+h$ et $x-h \leq \eta_2 \leq x$.

$$f''(x) \simeq \frac{f(x+h) + f(x-h) - 2f(x)}{h^2}$$

l'erreur est en $O(h^2)$.

Pour obtenir les formules de la troisième et la quatrième dérivée, on prend une combinaison linéaire des développement de Taylor, pour $f(x+2h), f(x+h), f(x-h)$ et $f(x-2h)$.

La table suivante donne différentes formules centrales toutes en $O(h^2)$:

$$
\begin{aligned}
f'(x_i) &\simeq \frac{1}{2h}\left[f(x_{i+1}) - f(x_{i-1})\right] \\
f''(x_i) &\simeq \frac{1}{h^2}\left[f(x_{i+1}) - 2f(x_i) + f(x_{i-1})\right] \\
f'''(x_i) &\simeq \frac{1}{2h^3}\left[f(x_{i+2}) - 2f(x_{i+1}) + 2f(x_{i-1}) - f(x_{i-2})\right] \\
f^{(4)}(x_i) &\simeq \frac{1}{h^4}\left[f(x_{i+2}) - 4f(x_{i+1}) + 6f(x_i) - 4f(x_{i-1}) + f(x_{i-2})\right].
\end{aligned}
$$

En utilisant les polynômes d'interpolation de Lagrange les dérivées d'ordre p sont calculées par:

$$f^{(p)}(\alpha) \sim \sum_{i=0}^{n} A_i(\alpha) f(x_i)$$

où

$$
\begin{aligned}
A_i(\alpha) &= L_i^{(p)}(\alpha) \quad p \leq n \\
\sum_{i=0}^{n} A_i(\alpha) x_i^k &= 0 \quad 0 \leq k \leq p-1 \\
\sum_{i=0}^{n} A_i(\alpha) x_i^k &= k(k-1)...(k-p+1)\alpha^{k-p} \qquad p \leq k \leq n.
\end{aligned}
$$

Remarque :

La fomule est exacte pour les polynômes de degrés $\leq n$.

Le système linéaire donnant les $A_i(\alpha)$ a un déterminant de type Vandermonde différent de zéro si les x_i sont distincts.

Les $A_i(\alpha)$ sont indépendants de f et peuvent être calculés une fois pour toutes.

5.2.4 Etude de l'erreur commise.

D'aprés le chapitre précédent, si f est connue en $(n+1)$ points $x_i, i = 0, ..., n$ alors $f(x) = P_n(x) + e(x)$, où $e(x)$ est l'erreur d'interpolation. En dérivant on obtient

$$\begin{aligned}
f'(x) &= P_n'(x) + e'(x) \\
&= \sum_{i=0}^{i=n} A_i(x).f(x_i) + e'(x) \\
\text{et} \quad e'(x) &= \frac{d}{dx}\left(\frac{1}{(n+1)!}L(x).f^{(n+1)}(\xi_x)\right) = \frac{d}{dx}\left(L(x).f[x_0, ..., x_n, x]\right) \\
&= \frac{1}{(n+1)!}L'(x).f^{(n+1)}(\xi_x) + \frac{1}{(n+1)!}L(x).\frac{d}{dx}\left(f^{(n+1)}(\xi_x)\right)
\end{aligned}$$

On remarque tout de suite que l'erreur de dérivation est nulle si f est un polynôme de degré inférieur ou égale à n. Si on prend pour x un point x_i, le second terme de la dérnière somme s'annule, sinon il faut connaître $\frac{d}{dx}\left(f^{(n+1)}(\xi_x)\right)$, ce qui est difficile car la fonction $x \to \xi_x$ étant inconnue. On peut donner une forme si f est $n+2$ fois dérivable en utilisant la notion de différence. En effet

$$\begin{aligned}
\frac{d}{dx}\left(f^{(n+1)}(\xi_x)\right) &= \frac{d}{dx}\left(f[x_0, ..., x_n, x]\right) \\
&= \lim_{h\to 0}\frac{f[x_0, ..., x_n, x+h] - f[x_0, ..., x_n, x]}{h} \\
&= \lim_{h\to 0} f[x_0, ..., x_n, x, x+h] \\
&= \lim_{h\to 0}\frac{1}{(n+2)!}f^{(n+2)}(\theta_{x,h}).
\end{aligned}$$

On constate qu'on devra se contenter d'une estimation

$$| e(x) | \leq \frac{1}{(n+1)!} | L'(x) | M_{n+1} + \frac{1}{(n+2)!} | L(x) | M_{n+2.}$$

5.3 Méthodes numériques d'intégration.

Le but de cette leçon est de calculer numériquement des intégrales définies ou indéfinies. Soit $f : [a, b] \to \mathbb{R}$, une fonction continue donnée. On désire approcher numériquement la quantité $\int_a^b f(x)dx$.

5.3.1 Formules fermées.

On appelle ainsi les formules quand la fonction f est continue sur lintervalle $[a, b]$. Les points d'interpolation x_i verifient $a = x_0 < x_1 < ... < x_{n-1} < x_n = b$.

Formule des rectangles.

La formule des rectangles est une formule dite à un point $x_0 = a$. Le polynôme d'interpolation associé est $P_0(x) = f(a)$ et $L(x) = x - a$ pour tout x appartenant à $[a, b]$. D'où

$$I(f) \simeq I(P_0) = f(a)(b - a).$$

L'interprétation graphique consiste donc à remplacer $\int_a^b f(x)dx$ par l'aire du rectangle de base $[a, b]$ et de hauteur $f(a)$.

Formule des trapèzes.

La formule des trapèzes est une formule à 2 points : $x_0 = a$ et $x_1 = b$. Le polynôme de Lagrange associé à ces deux points est $P_1(x) = f(a)\left(\frac{x-b}{a-b}\right) + f(b)\left(\frac{x-a}{b-a}\right)$.D'où

$$I(f) \simeq I(P_1) = \int_a^b P_1(x)dx = \frac{f(a) + f(b)}{2}(b - a).$$

Formule de Simpson.

La formule de Simpson est une formule à trois points $x_0 = a$, $x_1 = \frac{a+b}{2}$ et $x_2 = b$: . Le polynôme associé à ces trois points est $P_2(x) = f(a)L_0(x) + f(\frac{a+b}{2})L_1(x) + f(b)L_3(x)$. Notons que

$$L_0(x) = \frac{(x - x_1)(x - b)}{(a - x_1)(a - b)} \Rightarrow \int_a^b L_0(x)dx = \frac{(b - a)}{6},$$

$$L_1(x) = \frac{(x - a)(x - b)}{(x_1 - a)(x_1 - b)} \Rightarrow \int_a^b L_1(x)dx = \frac{4(b - a)}{6},$$

$$L_2(x) = \frac{(x - x_1)(x - a)}{(b - x_1)(b - a)} \Rightarrow \int_a^b L_2(x)dx = \frac{(b - a)}{6},$$

On tire donc la formule suivante:

$$I(f) \simeq I(P_2) = \frac{(b - a)}{6}\left[f(a) + 4f(\frac{a + b}{2}) + f(b)\right].$$

Formules ouvertes.

On appelle ainsi les formules quand la fonction f est continue sur l'intervalle $]a, b[$. Les points d'interpolation x_i verifient $a < x_0 < x_1 < ... < x_{n-1} < x_n < b$.

Formule de Steffensen.

Il en existe une infinité.

- Une à 1 points avec $x_0 = \frac{a+b}{2}$ qui donne la formule du milieu suivant:

$$I(f) \simeq (b-a)f(\frac{a+b}{2})$$

 Cette formule est exacte pour tout polynôme de degré 1.

- Une à 2 points avec $x_0 = \frac{2a+b}{3}$ et $x_1 = \frac{a+2b}{3}$ qui donne la formule suivante :

$$I(f) \simeq \frac{b-a}{2}\left(f\left(\frac{2a+b}{3}\right) + f\left(\frac{2b+a}{3}\right)\right).$$

 Cette formule est exacte pour tout polynôme de degré 1.

- Une à 3 points avec $x_0 = \frac{3a+b}{4}$ et $x_1 = \frac{a+b}{2}$ et $x_2 = \frac{3b+a}{4}$ qui donne la formule suivante :

$$I(f) \simeq \frac{b-a}{6}\left(4f\left(\frac{3a+b}{4}\right) + 2f\left(\frac{a+b}{2}\right) - 2f\left(\frac{a+3b}{4}\right)\right).$$

 Cette formule est exacte pour tous les polynômes de degré 2.

5.3.2 Etude générale de l'erreur commise.

Pour que les formules d'intégration numérique données précédement soient intéréssantes, il faut que l'on puisse estimer l'erreur $E(f) = I(f) - I(P_n)$ avec précision. Or si f est suffisamment dérivable, on a

$$E(f) = I(f - P_n) = \int_a^b \left[\frac{1}{(n+1)!} f^{(n+1)}(\xi_x) L(x)\right] dx.$$

Théorème : Supposons que $E(f) = 0$ pour les polynômes de degré au plus n et que la fonction $f \in C^{n+1}([a, b])$. On dit alors que la méthode est d'ordre

$n+1$. Si on pose $M_{n+1} = \max\limits_{x\in[a,b]} \mid f^{(n+1)}(x) \mid$, Une première estimation de l'erreur est

$$\mid E(f) \mid \leq \frac{1}{(n+1)!} M_{n+1} \int_a^b \mid L(x) \mid dx.$$

Théorème : En plus des hypothèses du Th précédent, on suppose que le polynôme $L(x)$ ne change pas de signe sur $[a,b]$, alors en utilisant le Th de la moyènne pour $E(f)$, on obtient

$$E(f) = \frac{1}{(n+1)!} f^{(n+1)}(\eta) \int_a^b L(x) dx.$$

$$\eta \in [a,b]$$

En utilisant ce dernier Théorème on peut estimer les erreurs des méthodes vues ci-dessus.

- **Pour la formule du rectangle on a:**

$$E(f) = f'(\eta) \int_a^b (x-a)dx = f'(\eta)\frac{(b-a)^2}{2} \qquad \eta \in [a,b]$$

  cette méthode est d'ordre 1.

- **Pour la formule du trapèze on a:**

$$E(f) = \frac{1}{2} f''(\eta) \int_a^b (x-a)(x-b)dx = -\frac{f''(\eta)}{12}(b-a)^3$$

  la méthode de Trapèze est dordre 2.

- **Pour la formule de Simpson on a:**

$$E(f) = -\frac{f^{(4)}(\eta)}{90} \left[\frac{b-a}{2}\right]^5,$$

  la méthode de Simpson est dordre 4.

Exemple :

$I = \int_0^1 e^{-x^2} dx, \quad a = 0, \frac{a+b}{2} = \frac{1}{2}, b = 1, f(0) = 1, f(\frac{1}{2}) = .7788, f(1) = .36788.$

1. Rectangle: $I \simeq f(0) = 1$.
2. Trapèze: $I \simeq \left[\frac{f(0)+f(1)}{2}\right] = .68393$.
3. Simpson: $I \simeq \frac{1}{6}\left[f(0) + 4f(\frac{1}{2}) + f(1)\right] = .74718$.
4. La valeur de I à 5 décimales est .74718.

5.3.3 Formules composées.

Plutôt que d'augmenter le degré du polynôme d'interpolation, on peut obtenir une formule d'integration en découpant l'intervalle d'intégration en sous-intervalles et en appliquant des formules simples sur chacun des sous-intervalles.

Formule de trapèze.

Si n est entier, posons

$$h = \frac{b-a}{n}, x_k = a + kh,\ \ k = 0, ..., n.$$

alors

$$\begin{aligned} I(f) = \int_a^b f(x)dx &= \sum_{k=0}^{n-1} \left(\int_{x_k}^{x_{k+1}} f(x)dx \right) \\ &= \sum_{k=0}^{n-1} \left[\left(\frac{f(x_k) + f(x_{k+1})}{2} \right) h - \frac{h^3}{12} f''(\eta_k) \right], \end{aligned}$$

où $\quad \eta_k \in [x_k, x_{k+1}], k = 0, ..., n-1$

Développant et regroupant les termes qui apparaissent 2 fois, on obtient

$$I(f) = \frac{h}{2}\left[f(a) + 2\sum_{k=1}^{n-1} f(a+kh) + f(b) \right] - \frac{h^3}{12}\sum_{k=0}^{n-1} f''(\eta_k)$$

En appliquant le Th des valeurs intermédiaires, on peut réécrire l'erreur sous la forme

$$E(f) = -\frac{nh^3}{12} f''(\eta) = -\frac{(b-a)}{12} f''(\eta) h^2.$$

Ceci nous donne la formule du trapèze composée pour laquelle l'approximation est donnée par:

$$T_n(f) = \frac{h}{2}\left[f(a) + 2\sum_{k=1}^{n-1} f(a+kh) + f(b)\right]$$

et l'erreur par

$$ET(f) = -\frac{(b-a)}{12} f''(\eta) h^2.$$

Formule de Simpson composée.

Supposons maintenant que n soit pair, groupant les intervalles 2 à 2 et appliquant la formule de Simpson sur $[x_i, x_{i+2}]$, on obtient

$$I(f) = \frac{h}{3}\left[f(a) + 4\sum_{k\ \ impair} f(a+kh) + 2\sum_{k\ \ pair} f(a+kh) + f(b)\right] - \frac{n}{2}\frac{f^{(4)}(\eta)}{90} h^5.$$

Ceci nous conduit à la formule de Simpson composée pour laquelle l'approximation est donnée par

$$S_n(f) = \frac{h}{3}\left[f(a) + 4\sum_{k\ \ impair} f(a+kh) + 2\sum_{k\ \ pair} f(a+kh) + f(b)\right]$$

et l'erreur par

$$ES(f) = -f^{(4)}(\eta)\frac{(b-a)}{180} h^4.$$

Exemple : Déterminer $\int_0^1 e^{-x^2} dx$.

Si n désigne le nombre des intervalles utilisés.

n	$T_n(f)$	$ET(f)$
2	.73137	.015
4	.74298	3.84×10^{-3}
8	.74658	9.58×10^{-4}
16	.74676	1.39×10^{-4}
32	.74680	5.98×10^{-5}

Si nous désirons obtenir 6 décimales exactes, il nous faut déterminer h tel que

$$\max_{0 \leq \eta \leq 1} \mid f''(\eta) \mid \frac{h^2}{12} \leq 5 \times 10^{-7}, \tag{5.1}$$

Pour une partition régulière $x_k = kh, h = \frac{1}{n}$; donc nous cherchons n tel que

$$n^2 \geq \frac{1}{12} \max_{0 \leq \eta \leq 1} \mid f''(\eta) \mid \frac{1}{5 \times 10^{-7}}.$$

or $f''(x) = e^{-x^2}(4x^2 - 2)$ et $f'''(x) = e^{-x^2} 4x(3 - 2x^2)$. Puisque $f'''(x)$ ne change pas de signe sur $[0; 1]$,

$$\max_{0 \leq \eta \leq 1} \mid f''(\eta) \mid = \max \{ \mid f''(0) \mid , \mid f''(1) \mid \} = 2.$$

On voit que (5.1) sera satisfaite si

$$n^2 \geq \frac{10^6}{3}, \qquad n > 578.$$

Remarque Dans le choix de la précision demandée, il faut tenir compte des erreurs d'arrondi et de l'accumutation des erreurs

5.4 Exercices

Série integration et dérivation.

Pour les problèmes des exercices (5.4) et (5.4), donner des approximations des dérivées dans les cas suivants:

En utilisant la formule de différence progressive.

En utilisant la formule de différence regressive.

En utilisant la formule de différence centrale.

Exercice : 1
Approcher $y'(1.0)$ si

$$x = [0.8 \quad 0.9 \quad 1.0 \quad 1.1 \quad 1.2]$$
$$y = [0.992 \quad 0.999 \quad 1.000 \quad 1.001 \quad 1.008]$$

Exercise: 2

1. Approcher $y'(4)$ si

$$x = [0 \quad 1 \quad 4 \quad 9 \quad 16]$$
$$y = [0 \quad 1 \quad 2 \quad 3 \quad 4]$$

2. Donner une expression de l'erreur de dérivation en $x = 4$.

3. Donner une majoration de l'erreur independament de x et de ξ_x.

Exercise : 3
Calculer $y''(2)$ si

$$x = [0 \quad 1 \quad 2 \quad 3 \quad 4]$$
$$y = [0 \quad 1 \quad 4 \quad 9 \quad 16]$$

Exercise : 4
Calculer $\int_0^{\frac{\pi}{2}} \sin^2 x dx$ en utilisant la formule du trapèze et la formule de Simpson. Comparer avec le résultat exact.

Exercise : 5
Pour le problème $P1$ approcher l' integrale:

1. En utilisant la formule de trapèze composée avec 2 intervalles.
2. En utilisant la formule de trapèze composée avec 10 intervalles.
3. En utilisant la formule de Simpson avec 2 intervalles.
4. En utilisant la formule de Simpson composée avec 10 intervalles.

$$P1 : \quad \int_0^1 x \sin(\pi x) dx$$

Exercise : 6

En utilisant les formules d'estimation d'erreur, trouver les bornes d'erreur pour le problème P1 dans les cas 1-4, puis calculer la valeur exacte de l'integrale et comparer les erreurs exactes "$E(f)$" et les bornes d'erreurs trouvées.

[1]

[1] S. El Bernoussi, S. El Hajji et A. Sayah